Universidade Federal do Rio de Janeiro
Instituto COPPEAD de Administração

O uso de ferramentas quantitativas em call centers – o caso Contax

Marco Aurélio Carino Bouzada

Doutorado em Administração

Orientador: Prof. Eduardo Saliby

Rio de Janeiro
2006

O uso de ferramentas quantitativas em call centers – o caso Contax

Marco Aurélio Carino Bouzada

Tese submetida ao corpo docente do Instituto COPPEAD de Administração (UFRJ) como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências (D. Sc.).

Aprovada por:

Prof. _____ - Presidente da Banca
Eduardo Saliby (COPPEAD/UFRJ)

Prof. _____
Alberto Gabbay Canen (COPPE/UFRJ)

Prof. _____
Edson Dalto (IBMEC)

Prof. _____
Kleber Fossati Figueiredo (COPPEAD/UFRJ)

Prof. _____
Ricardo Miyashita (UERJ)

Rio de Janeiro, RJ – Brasil

Dezembro de 2006

Ficha Catalográfica

Bouzada, Marco.

O uso de ferramentas quantitativas em call centers – o caso Contax / Marco Bouzada. Rio de Janeiro: COPPEAD, 2006.

xvi, 256p.

Tese – Universidade Federal do Rio de Janeiro, COPPEAD.

1. Call Centers (Administração). I. Tese (Doutor. – COPPEAD/UFRJ). II. Título.

A meus pais, Héctor e Lenise

Agradecimentos

Ao meu orientador, Eduardo Saliby,
pela inestimável ajuda no direcionamento desta pesquisa.

Aos demais integrantes da banca de defesa do meu projeto de tese,
professores Alberto Canen, Edson Dalto e Kleber Figueiredo,
pelos valiosos feedbacks proporcionados.

Aos membros da Gerência de Planejamento de Tráfego da Contax,
em especial aos srs. Alberto Porto, Alessandro Meira e Mario Pires,
pela presteza e qualidade na divulgação de informações.

À gerente comercial da Paragon, Erica,
pelo fornecimento da cópia do software Arena Contact Center.

À minha colega de linha de pesquisa, Juliana Yonamine,
pela discussão de idéias e troca de informações.

RESUMO DA TESE APRESENTADA AO COPPEAD/UFRJ COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS (D. SC.)

O USO DE FERRAMENTAS QUANTITATIVAS EM CALL CENTERS – O CASO CONTAX

MARCO AURÉLIO CARINO BOUZADA

2006

ORIENTADOR: PROF. EDUARDO SALIBY

PROGRAMA: ADMINISTRAÇÃO

Esta tese descreve – através do estudo de um caso – a aplicação de algumas ferramentas quantitativas a uma parte dos problemas operacionais encontrados no *call center* de uma grande empresa brasileira no setor – a Contax – e discute a sua adequabilidade.

Primeiramente, o problema estudado é introduzido através da explicitação dos objetivos da pesquisa – a respeito dos dois principais desafios atacados (a previsão de demanda e o dimensionamento da capacidade) e da sua importância para o meio acadêmico.

Depois de uma contextualização do setor e da empresa estudados, a literatura acerca dos principais temas abordados durante o estudo (métodos de previsão de demanda, teoria de filas e simulação, além de suas respectivas aplicações em *call centers*) é revista.

Os aspectos metodológicos são, então, apresentados, destacando o uso do estudo de um caso para obter maior profundidade na análise dos problemas e sugestão de soluções.

Na seqüência, o caso ambientado na empresa estudada é apresentado, com enfoque na descrição dos principais desafios operacionais e na indicação e no desenvolvimento de soluções quantitativas concretas para os mesmos.

Finalmente, os principais resultados obtidos são mostrados, com destaque para a adequação do uso da regressão múltipla ao problema de previsão de chamadas – tendo utilizado sistematicamente informações disponíveis capazes de influenciar a demanda – e do uso da simulação ao problema do dimensionamento da capacidade de atendimento – principalmente pela contemplação do comportamento de abandono dos clientes e de diferentes distribuições estatísticas para o tempo de atendimento, além de peculiaridades das operações da indústria em questão, responsáveis por corroborar a sua adequação.

ABSTRACT OF THESIS PRESENTED TO COPPEAD/UFRJ AS PARTIAL FULFILLMENT FOR THE DEGREE OF PHILOSOPHY DOCTOR (PH. D.)

O USO DE FERRAMENTAS QUANTITATIVAS EM CALL CENTERS – O CASO CONTAX

MARCO AURÉLIO CARINO BOUZADA

2006

CHAIRMAN: PROF. EDUARDO SALIBY

DEPARTMENT: ADMINISTRATION

This thesis describes – with the aid of a case study – the deployment of a set of quantitative tools to deal with part of the operation problems found in the call center of a large brazilian company in this industry called Contax, and discusses its usability.

The question is introduced with the description of the research objectives, regarding the two main challenges (demand forecast and capacity sizing) and its relevance within the academic context.

After presenting the industry and the company studied within it, the work shows the review of the literature regarding the main topics approached during the study: demand forecast methods, queuing theory, and simulation (including their use within call center industry).

The methodological issues are then presented, highlighting the use of the case study research in order to obtain a greater depth during the problems analysis and solutions suggestion.

The next part presents the case researched on the company, describing the main operational challenges and the concrete quantitative solutions identified and developed.

Finally, the main results are presented, including the appropriateness of the multiple regression method when dealing with the demand forecast problem (systematically using available data capable of influencing the demand) and of the simulation tool when dealing with the capacity sizing problem (contemplating the costumer abandonment behavior, the use of different statistical distributions for answering time and call centers' operational singularities able to grant the approach´s adequacy).

Lista de figuras

Figura 1 – Dimensões no atendimento ao cliente de *call center* 2
Figura 2 - Estrutura de relacionamento entre as entidades participantes na indústria de *Call Centers*..... 21
Figura 3 - Evolução do faturamento, do número de colaboradores e de posições de atendimento da Contax, 2000 a 2005..... 26
Figura 4 – Modelagem de um *call center* usando o Arena Contact Center 61
Figura 5 – Tela de entrada de variáveis do callLab 64
Figura 6 – Tela de resultados do callLab..... 65
Figura 7 – Tela de parâmetros do SimACD 66
Figura 8 – Tela de resultados do SimACD..... 67
Figura 9 – Relatório de saída do IVR Simulator Model..... 71
Figura 10 – Arquitetura do TotalView 88
Figura 11 – Tela de acompanhamento da quantidade de chamadas do TotalView.....89
Figura 12 – Tela de verificação da aderência do TotalView 90
Figura 13 – Tela de monitoramento do CentreVu Supervisor 91
Figura 14 – Relatório do CentreVu Supervisor 92
Figura 15 – Atividades da Gerência de Planejamento de Tráfego 93
Figura 16 – Organograma da Gerência de Planejamento da Contax..... 93
Figura 17 – Macro-processo de previsão de ligações e TMA.....101
Figura 18 – Minutagem por agente para cenários com diferentes efetivos de atendentes, produto 102, consolidado Brasil, abril de 2006..... 119
Figura 19 – Perfil histórico do comportamento intra-diário da demanda do produto 102, terça-feira 122
Figura 20 – Necessidade e disponibilidade consolidada (4 e 6 horas) de agentes por período, produto 102 (*plus*), Brasil, agosto de 2006 125
Figura 21 – Agentes escalonados e requeridos ao longo do dia, exemplo de escalonamento com cobertura total 126
Figura 22 – Agentes escalonados e requeridos ao longo do dia, exemplo de escalonamento balanceado 126
Figura 23 – Necessidade e disponibilidade (após o escalonamento) de agentes por período, produto 102 (*plus*), Brasil, agosto de 2006 131
Figura 24– Mudanças realizadas durante o escalonamento dos atendentes (6 horas), produto 102 (*plus*), Brasil, agosto de 2006..... 132
Figura 25 – Resumo do efetivo sugerido pelo dimensionamento para o mês subseqüente, produto 102 (*plus*), Brasil, agosto de 2006 133
Figura 26 – Quantidade de ligações por dia (multiplicada por uma constante), para o produto 103, julho de 2005 a março de 2006..... 146
Figura 27 – TMA por dia (multiplicado por uma constante), para o produto 103, julho de 2005 a março de 2006..... 146
Figura 28 – Comportamento sazonal do dia da semana – modelo modificado (sem V+3) para prever a quantidade de ligações, produto 103 154
Figura 29 – Impacto dos dias em torno do vencimento – modelo modificado (sem V+3) para prever a quantidade de ligações, produto 103 155
Figura 30 – Equação do modelo modificado (sem V+3) para prever a quantidade de ligações, produto 103 156
Figura 31 – Valores reais (multiplicados por uma constante) e previstos (pelo modelo) da quantidade de ligações por dia, para o produto 103, julho de 2005 a março de 2006..... 157
Figura 32 – Comportamento sazonal do dia da semana – modelo modificado (sem C; V – 3; V – 2) para prever o TMA, produto 103 167
Figura 33 – Impacto dos dias em torno do vencimento – modelo modificado (sem C; V – 3; V – 2) para prever o TMA, produto 103 168
Figura 34 – Equação do modelo modificado (sem C; V – 3; V – 2) para prever o TMA, produto 103 169
Figura 35 – Valores reais (multiplicados por uma constante) e previstos do TMA por dia, para o produto 103, julho de 2005 a março de 2006 170

Figura 36 – Diagrama de processos do modelo básico de simulação, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes 177
Figura 37 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 178
Figura 38 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 179
Figura 39 – Chamadas abandonadas e retornadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 180
Figura 40 – Tempos de atendimento e tempos de espera, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 181
Figura 41 – Agentes disponíveis e ocupados, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 182
Figura 42 – Nível de serviço para diferentes valores do parâmetro *k* da distribuição Erlang, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 184
Figura 43 – Taxa de abandono para diferentes valores do parâmetro *k* da distribuição Erlang, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 185
Figura 44 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 11 atendentes, 100 replicações 187
Figura 45 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 11 atendentes, 100 replicações 188
Figura 46 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 11 atendentes, 100 replicações 189
Figura 47 – Diagrama de processos do modelo de simulação com atendimento agregado (produto 102, básico e *plus*), das 00:00 às 00:30, agosto de 2006 193
Figura 48 – Chamadas geradas das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações 194
Figura 49 – Chamadas abandonadas e retornadas das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações 195
Figura 50 – Chamadas atendidas das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações 196
Figura 51 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações 197
Figura 52 – Tempos de atendimento e tempos de espera, das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações 198
Figura 53 – Utilização dos agentes "básicos" e "*plus*", das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações 199
Figura 54 – Utilização dos agentes "básicos" e "*plus*" por tipo de chamada, das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações 200
Figura 55 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% de demanda, 100 replicações 201
Figura 56 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% de demanda, 100 replicações 202
Figura 57 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% de demanda, 100 replicações 203
Figura 58 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% no TMA, 100 replicações 204
Figura 59 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% no TMA, 100 replicações 205
Figura 60 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% no TMA, 100 replicações 206
Figura 61 – Nível de serviço para diferentes valores do tempo médio de atendimento (TMA), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 208

Figura 62 – Taxa de abandono para diferentes valores do tempo médio de atendimento (TMA), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 209
Figura 63 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com tempo médio de espera antes do abandono = 1,5 minutos, 100 replicações 210
Figura 64 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com tempo médio de espera antes do abandono = 1,5 minutos, 100 replicações 211
Figura 65 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com tempo médio de espera antes do abandono = 1,5 minutos, 100 replicações 213
Figura 66 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com NS = % de clientes atendidos em até 5 segundos, 100 replicações 215
Figura 67 – Chamadas atendidas em até 5 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com NS = % de clientes atendidos em até 5 segundos, 100 replicações 216
Figura 68 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 13 atendentes e NS = % de clientes atendidos em até 5 segundos, 100 replicações 217
Figura 69 – Chamadas atendidas em até 5 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 13 atendentes e NS = % de clientes atendidos em até 5 segundos, 100 replicações 218
Figura 70 – Diagrama de processos do modelo de simulação com quantidade limitada (20) de linhas disponíveis, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes 219
Figura 71 – Indicadores dinâmicos do modelo de simulação com quantidade limitada (20) de linhas disponíveis, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes 220
Figura 72 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações 221
Figura 73 – Chamadas bloqueadas, abandonadas e retornadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações 222
Figura 74 – Chamadas completadas e atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações 223
Figura 75 – Disponibilidade, ocupação e utilização de linhas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações 224
Figura 76 – Chamadas geradas, atendidas e atendidas em até 10 segundos, das 05:30 às 06:00, agosto de 2006, produto 102 (*plus*), 11 atendentes, 100 replicações 226
Figura 77 – Chamadas geradas, atendidas e atendidas em até 10 segundos, das 02:00 às 02:30, agosto de 2006, produto 102 (*plus*), 4 atendentes, 100 replicações 228

Lista de tabelas

Tabela 1 – Magnitude atual e crescimento dos principais segmentos de atuação dos *call centers* no Brasil ... 20
Tabela 2 – Características gerais dos *softwares* de simulação em *call centers* pesquisados... 72
Tabela 3 – Atributos operacionais contemplados pelos *softwares* de simulação em *call centers* pesquisados ... 73
Tabela 4 – Magnitude do erro absoluto (%) médio por produto, 2006 ... 98
Tabela 5 – Pesos-padrão do TotalView usados no cálculo da média ponderada de demanda 102
Tabela 6 – Cronograma de envio e vencimento de contas do faturamento, maio de 2006..... 109
Tabela 7 – Cálculo de necessidade de recursos diários para o produto 102, central BA, cliente *plus*, outubro-novembro de 2006 ... 116
Tabela 8 – Premissas para a taxa de ocupação do atendente do produto 102, por dia da semana ... 116
Tabela 9 – Cálculo de necessidade de recursos diários para o produto 102, central BA, cliente *plus*, outubro-novembro de 2006 ... 118
Tabela 10 – Cálculo da necessidade de operadores para cada período de 30 minutos, produto 102 (*plus*), consolidado Brasil, agosto de 2006 ... 123
Tabela 11 – Necessidade (total) e disponibilidade (4 horas) de agentes por período, produto 102 (*plus*), consolidado Brasil, agosto de 2006 ... 124
Tabela 12 – Necessidade e disponibilidade de agentes e nível de serviço estimado por período, produto 102 (*plus*), Brasil, agosto de 2006 ... 128
Tabela 13 – Escala de atendentes e volume previsto de ligações para cada período de 30 minutos, produto 102 (*plus*), Brasil, novembro de 2005 ... 135
Tabela 14 – Nível de serviço estimado para cada período de 30 minutos, produto 102 (*plus*), Brasil, novembro de 2005 ... 136
Tabela 15 – Volume, TMA e nível de serviço estimados para os dias (hierarquizados) de novembro de 2005, produto 102 (*plus*), Brasil ... 137
Tabela 16 – Total de contas agrupadas por data de chegada prevista, setembro a outubro de 2005 ... 143
Tabela 17 – Total de contas agrupadas por data de vencimento, setembro a outubro de 2005 ... 144
Tabela 18 – Histórico da quantidade de ligações e do TMA (multiplicados por uma constante) para o produto 103, novembro de 2005 ... 145
Tabela 19 – Planilha de dados do modelo de regressão múltipla para prever a quantidade de ligações, produto 103, fevereiro de 2006 ... 148
Tabela 20 – Estatística de regressão do modelo de regressão múltipla para prever a quantidade de ligações, produto 103 ... 149
Tabela 21 – Análise de Variância do modelo de regressão múltipla para prever a quantidade de ligações, produto 103 ... 150
Tabela 22 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo de regressão múltipla para prever a quantidade de ligações, produto 103 ... 150
Tabela 23 – Estatística de regressão do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103 ... 151
Tabela 24 – Análise de Variância do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103 ... 152
Tabela 25 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103 ... 152
Tabela 26 – Valores reais, previstos e erros do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103, dezembro de 2005 ... 158
Tabela 27 – Valores reais, previstos pelo modelo e pelo TotalView e erros de previsão dos mesmos para a quantidade de ligações, produto 103, dezembro de 2005 ... 160
Tabela 28 – Planilha de dados do modelo de regressão múltipla para prever o TMA, produto 103, agosto de 2005 ... 162

Tabela 29 – Estatística de regressão do modelo de regressão múltipla para prever o TMA, produto 103 163
Tabela 30 – Análise de Variância do modelo de regressão múltipla para prever a o TMA, produto 103 163
Tabela 31 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo de regressão múltipla para prever o TMA, produto 103 164
Tabela 32 – Estatística de regressão do modelo modificado (sem C; V – 3; V – 2) de Regressão múltipla para prever o TMA, produto 103 165
Tabela 33 – Análise de Variância do modelo modificado (sem C; V – 3; V – 2) de regressão múltipla para prever o TMA, produto 103 165
Tabela 34 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo modificado (sem C; V – 3; V – 2) de regressão múltipla para prever o TMA, produto 103 166
Tabela 35 – Valores reais, previstos e erros do modelo (sem C; V – 3; V – 2) de regressão múltipla para prever o TMA, produto 103, janeiro de 2006 171
Tabela 36 – Valores reais, previstos pelo modelo e pelo TotalView e erros de previsão dos mesmos para o TMA, produto 103, janeiro de 2006 172
Tabela 37 – Nível de serviço e taxa de abandono para diferentes valores do parâmetro *k* da distribuição Erlang, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 183
Tabela 38 – Nível de serviço e taxa de abandono para diferentes quantidades de atendentes, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 100 replicações 190
Tabela 39 – Nível de serviço e taxa de abandono para diferentes valores do tempo médio de atendimento (TMA), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 207
Tabela 40 – Nível de serviço e taxa de abandono para diferentes valores do tempo médio de espera (antes do abandono), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações 214
Tabela 41 – NS real e estimado por fórmulas Erlang e simulação, erros de estimação e ganho de acurácia em diferentes patamares de nível de serviço 229
Tabela 42 – Nível de serviço e taxa de abandono para diferentes quantidades de atendentes, horário das 02:00 às 02:30, agosto de 2006, produto 102 (*plus*), 100 replicações 230

Lista de siglas

ABT	**Associação Brasileira de *Telemarketing***
ACD	***Automatic Call Distribution***
ANATEL	**Agência Nacional de Telecomunicações**
ANOVA	***Analysis of Variance***
CLT	**Consolidação das Leis do Trabalho**
CRM	***Costumer Relationship Management***
DAC	**Distribuidor Automático de Chamadas**
DMM	**Dia de Maior Movimento**
EP	**Erro Percentual**
EPSCC	***Excel Plataform for Simulating Call Centers***
FIFO	***First In, First Out***
FTE	***Full-Time Equivalent***
INMET	**Instituto Nacional de Meteorologia**
MAPE	***Mean Absolut Percentual Error***
NS	**Nível de Serviço**
PA	**Posição de Atendimento**
PEPS	**Primeiro a Entrar, Primeiro a Sair**
PSA	***Pointwise Stationary Approximation***
ROI	***Return on Investment***
SAC	**Serviço de Atendimento ao Cliente**
SLA	***Service Level Agreement***
SSA	***Simple Stationary Approximation***
SVD	***Singular Value Decomposition***
TMA	**Tempo Médio de Atendimento**
URA	**Unidade de Resposta Audível**

Sumário

1. Introdução.....1
1.1. O problema.....4
1.1.1. Objetivos.....4
1.1.2. Questões a serem respondidas.....4
1.1.2.1. Previsão.....5
1.1.2.2. Dimensionamento.....5
1.2. Relevância do estudo.....5
1.3. Delimitação do estudo.....12
1.4. Organização do estudo.....12
2. Caracterização do setor e da empresa estudados.....15
2.1. Setor de *Call Centers*.....15
2.1.1. *Call centers* no Brasil.....18
2.2. A Contax.....25
2.2.1. Histórico da empresa.....25
2.2.2. Números atuais da empresa.....27
2.2.3. Produtos e clientes da empresa.....27
3. Referencial teórico.....30
3.1. Métodos de previsão de demanda.....30
3.1.1. Séries temporais.....31
3.1.2. Séries causais.....34
3.1.2.1. Regressão linear múltipla.....34
3.2. Previsão de demanda de *call centers*.....35
3.3. Teoria de filas.....44
3.3.1. Teoria de filas aplicada a *call centers*.....47
3.4. Simulação.....50

3.5. Simulação em *call centers*.....52
3.5.1. *Softwares* de simulação para *call centers*.....59
3.5.1.1. Arena Contact Center.....60
3.5.1.2. callLAB.....63
3.5.1.3. SimACD.....65
3.5.1.4. ServiceModel.....67
3.5.1.5. Simul8.....68
3.5.1.6. ccProphet.....68
3.5.1.7. ContactCenters.....69
3.5.1.8. Excel Plataform for Simulating Call Centers (EPSCC).....69
3.5.1.9. IVR Simulator Model.....70
3.5.1.10. Quadros comparativos.....70
3.6. Resumo.....74
4. Metodologia.....78
4.1. Tipo de pesquisa.....78
4.2. Seleção do caso e dos sujeitos.....79
4.3. Coleta e tratamento dos dados.....81
4.3.1. Previsão.....81
4.3.2. Dimensionamento.....82
4.4. Limitações do método.....84
5. Descrição do caso.....88
5.1. Gerência de Planejamento de Tráfego da Contax.....88
5.2. Principais desafios operacionais.....95
5.2.1. Previsão de demanda.....96
5.2.2. Dimensionamento da capacidade de atendimento.....111
6. Análise do caso.....140
6.1. Previsão de demanda.....140
6.1.1. A escolha do método de previsão e do “produto-cobaia”.....141
6.1.2. O modelo de previsão para a demanda de ligações.....147
6.1.3. O modelo de previsão para o tempo médio de atendimento (TMA).....161
6.2. Dimensionamento da capacidade de atendimento.....173

7. Conclusões.....231
7.1. Respostas às questões propostas.....232
7.1.1. Previsão.....232
7.1.2. Dimensionamento.....233
7.2. Sumário das contribuições da pesquisa.....239
7.3. Sugestões e recomendações.....239
7.3.1. Ferramentas de previsão de demanda.....240
7.3.2. Dimensionamento da capacidade de atendimento.....241
7.3.3. Operador multi-produto.....244
8. Referências bibliográficas.....246
ANEXO I: *Checklist* da equipe de planejamento para auxiliar a elaboração da previsão de demanda.....256

1. Introdução

Um *call center*, também conhecido como central de atendimento, consiste em qualquer grupo cuja principal atividade de negócio seja falar com clientes (potenciais ou atuais) ao telefone com o objetivo de apoiar processos de vendas, *marketing*, serviços ao cliente, suporte técnico, ou outra atividade específica, funcionando como um elo entre a empresa e a sua base de clientes. (GROSSMAN *et al*, 2001; WEINBERG, BROWN & STROUD, 2006; SAKAMOTO, 2001; AZEVEDO & CALDAS, 2002)

Segundo Mehrotra, Profozich e Bapat (1997), os gerentes e planejadores dos *call centers* têm um emprego muito mais difícil hoje do que no passado. Com muito mais produtos e serviços sendo especialmente criados, disponibilizados no mercado, vendidos e assistidos, em relação a momentos anteriores, os *call centers* têm despendido muito esforço para fornecer diferentes níveis de serviço para diferentes tipos de clientes com diferentes necessidades. Os sistemas de telefonia atuais permitem grande flexibilidade e podem determinar como as chamadas devem ser roteadas e enfileiradas; mas, ao mesmo tempo, tornam a previsão, o planejamento e a análise ainda mais difíceis, pelo fato de tornarem possível o *link* entre múltiplos *call centers*, a priorização de certas chamadas, a existência de diferentes habilidades entre os operadores e a customização da lógica de roteamento das chamadas.

Hoje, os gerentes precisam estar aptos a entender o que está havendo nos *call centers*, para saber como as chamadas, rotas, prioridades, operadores e suas habilidades, a previsão de demanda e da carga de trabalho, terceirização, períodos de pico e outros fatores influenciam o nível de serviço, as taxas de abandono e de utilização.

De acordo com Hall e Anton (1998), um dos principais desafios atuais para os *call centers* consiste em fornecer serviço de valor agregado ao cliente ao menor custo possível por contato telefônico. Dessa forma, duas dimensões podem ser medidas: a eficiência do negócio e a efetividade para o cliente (conforme pode ser visto na figura 1 a seguir).

Figura 1 – Dimensões no atendimento ao cliente de *call center*

Satisfação do cliente		
Alta	**Efetivo** **Ineficiente**	***World-class*** **Eficiente e efetivo**
Baixa	**Péssimo** **Ineficiente e inefetivo**	**Eficiente** **Inefetivo**

Alto **Custo** Baixo

Fonte: HALL & ANTON (1998)

O gerente de um *call center* pode usar o modelo da figura 1 anterior para determinar o quadrante no qual a sua central está operando. Muitos *call centers* se encontram atualmente no quadrante inferior esquerdo da matriz (50% dos examinados em uma pesquisa da Universidade de Purdue), tendo, assim, à sua disposição, algumas ferramentas no mercado para ajudar no aumento da eficiência e da efetividade. Para passar para o quadrante superior direito, é necessária uma grande reengenharia dos processos.

Ainda segundo os autores, esses *call centers* podem usar uma ferramenta conhecida como Simulação para testar (e, eventualmente, justificar a implantação) se determinadas mudanças serão capazes de melhorar o sistema, antes de implantá-las. Os *call centers* "*world-class*" usam essa ferramenta efetiva e eficientemente para projetar o sistema, gerenciar a operação e se planejar para o futuro, diante de possíveis cenários.

Isso ocorre porque, entre outros motivos, o dimensionamento da capacidade de atendimento consiste em uma atividade crítica para a obtenção da eficiência e da efetividade da operação. E a ferramenta de simulação costuma se adequar melhor ao dimensionamento de operações mais complexas (como as de um *call center* moderno), já que ela consegue modelar muito bem a realidade, apresentando resultados mais acurados e relativamente precisos.

É bem verdade que esses resultados não são tão precisos quanto os resultados teóricos (obtidos por métodos analíticos), mas representam, normalmente, uma boa aproximação para os mesmos. A máxima da simulação diz que "é melhor ter uma solução aproximada para um modelo bem realista do que uma solução exata para um modelo com muitas aproximações".

Pois uma das principais vantagens da simulação em relação aos métodos analíticos consiste justamente na modelagem mais acurada da realidade: os métodos analíticos precisam "fazer muitos ajustes na realidade" para que esta se encaixe em algum padrão pré-concebido (exemplo: os tempos de atendimento devem seguir uma distribuição exponencial, o que, muitas vezes, está longe de corresponder à realidade) que a teoria analítica seja capaz de tratar, enquanto a simulação é capaz de representar muito bem quase qualquer peculiaridade de uma operação complexa (por exemplo: roteamento das chamadas, diversidade de funções dos atendentes, abandono de chamadas e outras características comuns nos *call centers* e que, dificilmente, poderiam ser contempladas por uma abordagem analítica). Naturalmente, essa vantagem cresce juntamente com a complexidade da operação a ser estudada e praticamente inexiste em operações extremamente simples.

Pode parecer contraditório dizer que a simulação consiste em uma abordagem menos precisa e mais acurada, quando comparada aos métodos analíticos. Mas, na verdade, é isso mesmo que ocorre: como os resultados são empíricos, eles variam a cada vez que a simulação é executada, gerando um pouco de imprecisão; entretanto, como estes resultados se baseiam em um modelo mais próximo da realidade, eles são realmente mais acurados (mais próximos dos "valores corretos" para aquele problema real).

Obviamente, a abordagem por simulação apresenta – em relação às abordagens analíticas – algumas desvantagens, além da menor precisão (citada anteriormente), destacando-se, entre outras: a necessidade de *know-how* e *software* específicos e o consumo de tempo para a modelagem do problema e para a simulação da operação em si. Estas desvantagens tornam a simulação uma ferramenta mais cara e de implementação mais difícil.

Além disso, uma ferramenta mais acurada e adequada, capaz de melhorar a qualidade do processo (de dimensionamento), trará poucos benefícios para o produto final deste processo (a quantidade necessária de atendentes) se o seu insumo apresentar uma qualidade insatisfatória.

A matéria-prima para o processo de dimensionamento em um *call center* consiste na previsão da demanda de ligações e da conseqüente carga de trabalho, que precisam, então, ser estimados de forma bem acurada. Essa necessidade confere também à previsão de demanda o *status* de atividade crítica durante a busca pela eficiência e pela efetividade do negócio.

1.1. O problema

As questões que emergem, então, são as seguintes:

Quais metodologias são usadas para lidar com o problema de dimensionamento de *call centers*? Será que este tipo de operação apresenta características e peculiaridades que justifiquem o uso de uma abordagem mais cara e de implementação mais difícil, como a simulação? Em caso afirmativo, como a simulação pode ajudar os gerentes das centrais no tratamento dos seus desafios operacionais?

Em relação à outra atividade crítica a ser também tratada quantitativamente – a previsão da demanda de chamadas – a mesma curiosidade científica aguçada na questão do dimensionamento motiva a indagação acerca de quais metodologias são utilizadas para resolver esse problema e se são elas as mais adequadas para tal fim.

1.1.1. Objetivos

- Levantar o que vem sendo feito para lidar com alguns dos principais desafios e problemas quantitativos relacionados à administração operacional no *call center* estudado (a ser especificado na seção 1.3 a seguir);
- Verificar a adequação de tais metodologias de resolução destes problemas;
- Sugerir metodologias mais adequadas (se for o caso) para a resolução destes problemas.

1.1.2. Questões a serem respondidas

A pesquisa envolve questões relacionadas a dois diferentes tipos de problemas dos *call centers*:

1. ao problema da **previsão** da quantidade de chamadas recebidas por unidade de tempo e do tempo médio de atendimento, *no call center* estudado;
2. ao problema do **dimensionamento** (contínuo durante a operação, em face de mudanças em algumas premissas e/ou parâmetros) da quantidade necessária de atendentes por horário, *no call center* estudado.

1.1.2.1. Previsão

a) Que metodologia(s) é(são) aplicada(s)?

b) Essa(s) é(são) a melhor forma de prever tais grandezas?

c) Se não for(em), qual é?

1.1.2.2. Dimensionamento

d) Qual metodologia é aplicada para a resolução desse problema?

e) A simulação é utilizada?

f) Se não for, por que não é?

g) Quais são os *trade-offs* envolvidos na decisão de qual metodologia usar (simulação *versus* outras)?

h) Como a simulação pode dar apoio às decisões referentes ao processo de dimensionamento do *call center*?

i) Que aspectos operacionais do *call center* não podem ser bem tratados por outras metodologias, sendo determinantes para justificar o uso da simulação?

j) A simulação fornece resultados mais acurados em algumas situações do cotidiano operacional do *call center*? Em quais situações?

k) Quais *softwares* estão disponíveis para o uso da simulação aplicada a *call centers*?

l) Quais são as suas características?

m) Quais deles são utilizados no *call center* estudado?

1.2. Relevância do estudo

Dezenas de bilhões de dólares foram gastos em *call centers* na última metade da década de 1990. O crescimento dessa indústria naquela década foi na faixa de 20% ao ano; e é esperado que esse crescimento se mantenha neste patamar durante o início desse século, segundo Grossman *et al* (2001).

Essa indústria cresceu dramaticamente na década em questão, empregando estimados 3% da mão-de-obra norte-amerciana. No início do século XXI, em torno de 10 milhões de pessoas estavam empregadas em *call centers*, somente nos EUA! Nunca os *call centers* constituíram uma parte tão importante da maioria dos negócios quanto hoje. No caso particular de vendas e serviços, *call centers* são freqüentemente o ponto de contato entre a empresa e os clientes, o que os torna críticos e caros. Empresas que não tomam as providências necessárias para melhorar a eficiência de seus *call centers* rapidamente descobrem que seus erros de planejamento são traduzidos em baixos níveis de serviço, vendas perdidas, aumentos nos custos e clientes frustrados e insatisfeitos. (BATT, DOELLGAST & KWON, 2004)

Na economia de hoje, os *call centers* não apenas se tornaram os pontos primários de contato entre clientes e empresas, mas também um grande investimento para muitas organizações. (WEINBERG, BROWN & STROUD, 2006)

Os gerentes atuais precisam entender o que está havendo nos *call centers*, para saber como as suas características influenciam os indicadores de desempenho. Utilizar palpites, tentativa-e-erro, intuição ou *softwares* "caixa-preta" (que ninguém na empresa é capaz de entender) é muito perigoso para empresas que querem ser bem sucedidas; e muito arriscado – em termos de estabilidade profissional – para os gerentes das centrais.

Chassioti e Worthington (2004) afirmam que existem diversos desafios gerenciais para as centrais de atendimento: recrutamento, absenteísmo, políticas de monitoramento das ligações, integração de sistemas, treinamento, entre outros. À medida que esses desafios são gerenciados de uma melhor forma e as centrais vão crescendo (inclusive em termos de custos), aumenta o interesse no uso de técnicas de pesquisa operacional pela indústria em questão.

Segundo Alam (2002) e Hall e Anton (1998), o custo de pessoal representa aproximadamente 70% do total da indústria, justificando a importância de um gerenciamento quantitativo da capacidade de atendimento, que consiste no *trade-off* entre esse custo e a determinação do nível ótimo de serviço; ou seja, em ter o número certo de pessoas qualificadas e recursos de suporte no momento correto, para lidar com uma previsão acurada da demanda e da carga de trabalho, mantendo os padrões de qualidade e o nível de serviço exigido. Por isso e à medida que os *call centers* têm apresentado operações cada vez mais complexas, tem se tornado ainda mais importante o uso de modelos acurados para a previsão de demanda e para o

dimensionamento do tamanho da equipe em uma indústria que movimenta tanto volume financeiro.

Segundo Mehrotra e Fama (2003) e Hall e Anton (1998), *call centers* são interessantes objetos de estudo da simulação, pois:

- lidam com mais de um tipo de chamada, onde cada tipo representa uma fila;
- as chamadas recebidas em cada fila chegam aleatoriamente ao longo do tempo;
- em alguns casos, agentes fazem ligações pró-ativamente (tipicamente em *telemarketing* ou ações de cobranças) ou como retorno de uma chamada anterior;
- a duração de cada chamada é aleatória, assim como a do trabalho que o agente executa depois da chamada (por exemplo: entrada de dados, documentação, pesquisa);
- o avanço nos sistemas para rotear as chamadas para os agentes, grupos ou localizações torna a lógica por trás do *call center* cada vez mais sofisticada;
- agentes podem ser treinados para responder a um tipo único de chamada, vários tipos de chamadas ou todos os tipos de chamadas, com diferentes prioridades e preferências especificadas nas lógicas de roteamento;
- a grande quantidade de dinheiro investido em *call centers*, tanto na forma de capital como na forma de trabalho, é capaz de justificar o uso dessa ferramenta tão poderosa.

Com os objetivos de esclarecer a questão e de mostrar as limitações de métodos analíticos para lidar com as peculiaridades reais e de difícil tratamento dos *call centers* (caracterizando as lacunas a serem preenchidas pela abordagem da simulação), um exemplo hipotético é descrito a seguir:

Um determinado *call center* recebe chamadas telefônicas que chegam à central com um intervalo entre chegadas que segue uma distribuição exponencial. A duração de cada atendimento também segue uma distribuição exponencial. A empresa quer dimensionar a central, determinando quantos atendentes deve alocar para cada horário. Para tal, ela quer construir um quadro comparativo, relacionando cada possível quantidade de atendentes com o nível de serviço prestado (na forma de tempo médio do cliente na fila, tempo médio do cliente

no sistema e tamanho médio da fila). Para essa operação, os modelos analíticos Erlang (**M/M/s**) (intervalo entre chegadas seguindo uma distribuição exponencial / tempo de atendimento seguindo uma distribuição exponencial / **s** agentes no sistema) atendem perfeitamente a demanda e fornecem fórmulas prontas para as variáveis de desempenho supracitadas em função do número de agentes (**s**), do intervalo médio entre chamadas e do tempo médio de atendimento. Dessa forma, o gerente desse *call center* pode dimensionar facilmente a sua equipe em cada horário, utilizando essa abordagem analítica e considerando o *trade-off* envolvendo o custo da mão-de-obra e o nível de serviço.

No entanto, se a esse exemplo forem adicionados alguns aspectos mais reais e que ocorrem freqüentemente no cotidiano operacional dos *call centers*, o panorama muda completamente. Pode acontecer de uma central querer desviar a chamada para um atendimento mais especializado, caso a ligação demore mais do que 5 minutos. Talvez o cliente que esteja aguardando para ser atendido desista após 3 minutos de espera (isso pode ocorrer sempre, ou de maneira incerta, com, por exemplo, 80% de probabilidade). É possível (e até provável) que a distribuição dos tempos de atendimento não seja exponencial; pode ser que esta variável não siga nenhuma distribuição teórica e tenha que ser tratada de acordo com uma distribuição empírica (exemplo: 20% de probabilidade de o atendimento demorar 2 minutos; 50% de probabilidade de o atendimento demorar 3 minutos; 30% de probabilidade de o atendimento demorar 4 minutos). Às vezes, a empresa quer priorizar clientes especiais, permitindo que estes sejam atendidos antes de outros clientes, mesmo que tenham chegado depois ao sistema.

Com a presença de algum(ns) desses complicadores – que são bem freqüentes no dia-a-dia das centrais de atendimento – a abordagem analítica torna-se praticamente inviável. Não existem modelos teóricos que forneçam fórmulas prontas (como no caso mais simples, hipotetizado anteriormente) para ajudar no dimensionamento da capacidade. Se o gerente quiser usar esses modelos analíticos, terá que fazer tantos ajustes à realidade (para que ela possa se encaixar nos modelos pré-concebidos que essas abordagens exigem) que, na prática, estará resolvendo um outro problema. Nesse caso, ele obterá uma solução exata para um modelo com muitas aproximações.

A simulação pode, entretanto, incorporar na modelagem todas essas características e dificuldades dos *call centers*, até com uma certa facilidade. Como o próprio nome sugere, a abordagem vai simular a operação do *call center* com os seus próprios aspectos, onde é

possível introduzir toda essa inteligência (orientação dizendo que, por exemplo, se a chamada demorar mais do que um determinado limite de tempo para ser atendida, algo deve acontecer, com um determinado percentual de probabilidade; ou que, se o atendimento durar mais do que um determinado limite de tempo, algum procedimento deve ocorrer), além das outras características, como distribuições empíricas dos tempos entre chamadas e/ou de atendimento, diferenciação de clientes, priorização de chamadas, dentre muitas possíveis.

Portanto, de uma maneira geral, os problemas básicos que aparecem nos *call centers* (e em outras estações de serviço) podem ser atacados através de uma metodologia de caráter analítico – a Teoria de Filas. Para ser utilizada esta abordagem, no entanto, é preciso engessar a modelagem do problema, deixando a operação de acordo com as restrições necessárias para que possa ser tratada através dessa teoria. Assim, quando as características da operação em questão fogem um pouco dos requisitos que a sua modelagem deve preencher para que ela possa ser tratada por essa metodologia, a teoria de filas deixa um pouco a desejar, já que a simplificação exigida pode ser muito exagerada, descaracterizando o aspecto real do problema.

É o pensamento de Mehrotra e Fama (2003), que pode ser evidenciado quando estes destacam algumas questões que – dado o aumento de complexidade de operação dos *call centers* – se tornaram difíceis de serem atacadas pelos modelos analíticos, passando a ser tratadas de forma mais adequada pela ferramenta de simulação:

- Quantos agentes devem ser contratados? Como devem ser escalados em relação aos turnos, lanches, treinamento e outras atividades?

- Como as chamadas devem ser roteadas de forma que se possa fazer o melhor uso dos recursos?

- Dada a previsão, o sistema de roteamento, a escala dos agentes, como será a performance do sistema?

- Qual o efeito de um pico de volume de chamadas na performance?

Enfim, é preciso descobrir se é efetivamente adequada a aplicação de uma ferramenta consagrada, tão poderosa e em voga no meio científico, como a simulação, a problemas de uma indústria tão complexa e que vem crescendo tanto ultimamente, como a de *call centers*.

Mais especificamente no Brasil, essa indústria tem crescido bastante. Um estudo recente (BRASIL fecha 2003 com mais de 68 mil Pas, 2005) apurou que o segmento das centrais de atendimento tornou-se um dos setores da economia brasileira que mais gerou empregos no país em 2003. Com esse mercado doméstico em efervescência, é preciso verificar se a simulação vem sendo amplamente utilizada no país.

O que ocorre, no entanto e de acordo com CAPES (2006), é que apenas 27 teses e dissertações brasileiras versam sobre o tema *Call Center*, segundo consulta utilizando as expressões-chave "*call center*" e "central de atendimento". Analisando mais profundamente esta lista, constata-se que apenas 3 destas 27 obras tratam do gerenciamento do processo operacional das centrais: Pazeto (2000); Silva Jr. (2004); e Moura (2004). Nenhuma delas, entretanto, aborda a ferramenta de simulação como forma de auxiliar o gerenciamento do processo operacional e o dimensionamento de capacidade das centrais de atendimento, o que também pode ser observado se a consulta à base incluir a palavra-chave "simulação", além das expressões-chave citadas anteriormente.

Esta é uma das principais contribuições deste trabalho: no Brasil, muito pouco foi produzido academicamente no que diz respeito à aplicação da ferramenta em questão (simulação) à indústria de *call centers*; em nível de teses e dissertações, esta produção acadêmica é incipiente.

Até em termos mundiais, a produção intelectual que versa sobre o tema (aplicação da ferramenta à indústria) é insatisfatória: Chassioti e Worthington (2004) afirmam que muitas empresas desenvolvem pesquisas particulares (não publicadas) sobre o assunto, mas é necessário mais trabalho de domínio publico acerca do tema.

A investigação a respeito de quais metodologias quantitativas têm sido utilizadas e a verificação da sua adequação fazem-se necessárias também em um outro problema operacional, quantitativo e de caráter crítico que os gerentes dos *call centers* têm que resolver freqüentemente: a previsão de demanda (chamadas recebidas por unidade de tempo).

Segundo Weinberg, Brown e Stroud (2006), devido à magnitude das operações dos *call centers*, seus supervisores precisam dimensionar eficientemente as suas equipes para fornecer satisfatórios níveis de serviço a custos razoáveis. A administração eficiente das centrais requer a estimação de algumas premissas operacionais que, em última análise, são capazes de

determinar tamanhos apropriados de equipe de atendimento. A previsão acurada da quantidade de ligações entrantes consiste na premissa mais difícil de ser obtida.

A base de qualquer bom planejamento de equipe é uma previsão acurada da carga de trabalho. Sem uma previsão precisa acerca do trabalho a ser esperado, os mais sofisticados esforços para calcular o contingente ótimo de atendimento e criar intrincados planos de horários consistem em desperdício. A máxima "*garbage in, garbage out*" (quando o que entra é "lixo", o que sai também é "lixo") é especialmente verdadeira quando aplicada ao gerenciamento da força de trabalho. Uma previsão acurada é o passo mais importante do processo. (REYNOLDS, 2005)

De acordo com Gans, Koole e Mandelbaum (2003), uma previsão acurada dos parâmetros do sistema consiste em um pré-requisito para um nível de serviço consistente e para uma operação eficiente. Dado o avanço tecnológico recente na área de informática e o ambiente intensivo em informações presente nos *call centers* modernos, poderia se supor que existiriam métodos de estimação e previsão bem desenvolvidos na literatura. Mas, de fato, apesar da vasta literatura em inferência estatística e previsão, surpreendentemente pouco tem sido dedicado aos processos estocásticos, e muito menos a modelos de filas, em geral, e a *call centers*, em particular.

Apenas 17 trabalhos científicos versam sobre estatísticas e previsão de informações de *call centers*, dentre os mais de 250 listados na bibliografia desenvolvida por Mandelbaum (2003), relacionando artigos ligados a *call centers*, nas mais diversas disciplinas.

Somente **um** item foi encontrado em uma pesquisa realizada em uma base de dados consagrada, procurando os termos "*call center*" (ou "*call centers*") **e** "*forecast*" (previsão) em todos os campos de busca (excluindo o texto completo) (*all fields – excluding full text*). (EMERALD, 2006).

Especificamente na indústria em questão, o estado-da-arte em previsão do volume de chamadas ainda é rudimentar. Na verdade, a prática de estatística e análise de séries temporais ainda está na sua infância no mundo dos *call centers*, e esforços de pesquisa sérios são necessários para trazê-las ao nível apropriado, dentro deste contexto. A previsão das taxas de chegada talvez consista na necessidade prática com maior pressão por melhorias. Para *call centers* com altos níveis de utilização, previsões mais acuradas são essenciais. Apesar de

existir alguma pesquisa que desenvolve métodos de estimação e predição das taxas de chegada, certamente existe espaço para melhorias adicionais. No entanto, o desenvolvimento destes métodos depende, em parte, do acesso a conjuntos de informação mais ricos. (GANS, KOOLE & MANDELBAUM, 2003)

1.3. Delimitação do estudo

A pesquisa se limita a estudar uma única grande empresa brasileira de *call centers*, a Contax, que será apresentada e contextualizada na seção 2.2 a seguir.

O estudo se restringe a algumas das operações de atendimento passivas (ou *inbound*) da empresa, onde seus operadores recebem ligações dos clientes e não as originam; operações do tipo *outbound* estão fora do escopo da pesquisa. O ponto de vista da análise reflete a ótica da empresa fornecedora do serviço de atendimento telefônico (e não a do cliente final ou da empresa contratante do serviço).

A idéia do trabalho consiste em abordar apenas questões quantitativas referentes à operação das centrais de atendimento, sem envolver outras questões igualmente relevantes e freqüentemente encontradas na literatura, como gestão de recursos humanos, ergonomia, terceirização e tecnologia.

As questões operacionais e quantitativas do *call center* estão restritas à previsão da quantidade de chamadas por unidade de tempo e do tempo médio de atendimento e ao dimensionamento da quantidade necessária de atendentes por horário, não englobando outras questões quantitativas também importantes, como a macro-localização física da instalação, entre outras.

1.4. Organização do estudo

A presente pesquisa foi estruturada em 7 capítulos:

1. Introdução
2. Caracterização do setor e da empresa estudados
3. Referencial teórico

4. Metodologia

5. Descrição do caso

6. Análise do caso

7. Conclusões

Este capítulo 1 buscou apresentar o tema com as suas principais questões pertinentes, definir o problema estudado (incluindo os objetivos e as questões a serem respondidas), argumentar a favor da relevância da pesquisa e especificar o seu escopo.

O capítulo 2 visa contextualizar o setor e a empresa estudados, apresentando a indústria de *call centers* (de uma forma geral e no Brasil) e descrevendo a empresa estudada: seu histórico, seus números atuais e seus produtos e clientes.

O capítulo 3 se dedica à revisão de literatura acerca dos principais assuntos abordados durante o estudo: métodos de previsão de demanda, teoria de filas e simulação, além de suas respectivas aplicações em *call centers*.

O capítulo 4 classifica a pesquisa quanto ao seu tipo e explica como e por que motivo o caso estudado foi selecionado, assim como o foram os principais sujeitos envolvidos. Em seguida, uma atenção especial é dedicada à descrição de como os dados foram coletados e tratados, de forma a dar resposta às questões e atingir os objetivos propostos. O capítulo é finalizado com as limitações intrínsecas do método e também oriundas da aplicação prática do mesmo.

O capítulo 5 descreve a Gerência de Planejamento de Tráfego da empresa estudada no caso, setor responsável pelo gerenciamento dos principais problemas atacados nesta pesquisa e apresenta os seus desafios operacionais mais importantes, mostrando também a forma como alguns deles são tratados pela empresa.

O capítulo 6 explora os principais desafios operacionais levantados no capítulo anterior – a previsão de demanda e o dimensionamento da capacidade de atendimento – questionando as abordagens utilizadas pela empresa para resolvê-los e apresentando soluções concretas para os mesmos, sempre de forma a fornecer resposta para as questões e atingir os objetivos, propostos inicialmente.

Finalmente, o capitulo 7 conclui o estudo, apresentando os principais resultados e contribuições obtidos e oferecendo ao leitor a versão do autor face à resposta ao problema e às suas questões. E, como é inviável operacionalmente tratar todos os assuntos aos quais está ligado o tema da pesquisa, o capítulo e o trabalho são finalizados com a sugestão de alguns tópicos correlatos que possam vir a constituir-se dos objetos de estudo de outros pesquisadores em trabalhos futuros.

2. Caracterização do setor e da empresa estudados

2.1. Setor de *Call Centers*

Sakamoto (2001) define *call center* como sendo "o lugar onde os chamados são feitos, ou recebidos, em grandes quantidades, com o objetivo de apoiar processos de vendas, *marketing*, serviços ao cliente, suporte técnico ou outra atividade específica".

Call centers consistem em centros operacionais instalados para utilizar telecomunicação e tecnologias computacionais para automatizar uma quantidade volumosa de diferentes atividades e serviços telefônicos, tanto recebidos quanto originados pelo centro. (HAWKINS *et al*, 2001)

Para Minghelli (2002), um *call center* pode ser entendido como um "centro integrado de contato entre empresas e consumidores, estabelecidos de forma remota e/ou virtual, através do uso da tecnologia".

Mais tecnicamente, um *call center* pode ser definido como um complexo sistema de componentes conectados, contemplando diferentes tipos de chamadas telefônicas, padrões de chegada de chamadas, lógica de roteamento das chamadas, abandono de chamadas, troncos de linhas telefônicas, e operadores com diferentes habilidades, prioridades, níveis de proficiência e esquemas de horários.

Do ponto de vista operacional, um *call center* caracteriza-se, basicamente, por um conjunto de tele-operadores, reunidos em um espaço e interligados ao público de uma empresa através de sistemas de telefonia, recebendo e realizando ligações para atender, tirar dúvidas, ouvir reclamações, resolver problemas ou vender produtos e serviços. Do ponto de vista do negócio, um *call center* é um meio de relacionamento entre empresas e seus clientes, que tem como objetivos: estreitar esse relacionamento; melhorar, acelerar e desburocratizar processos; promover a fidelização do cliente com a empresa por meio das melhorias no atendimento; e aumentar receitas. (WEISSHUHN, 2004).

De acordo com Weinberg, Brown e Stroud (2006), um *call center* consiste em um eixo centralizado que existe unicamente com o propósito de fazer (atender) chamadas para (de) os clientes (atuais ou potenciais).

Bapat e Pruitte Jr. (1998) argumentam que o *call center* é a arma estratégica mais visível de uma empresa. É uma frente de batalha de negócios onde milhões de dólares em produtos e serviços são comprados, vendidos e negociados. Também é um local onde milhares de clientes são conquistados e perdidos instantaneamente. Como empresas líderes se tornam mais criativas em disseminar informação e agregar valor aos seus consumidores através de linhas telefônicas, é natural que elas enxerguem o *call center* como seu ponto de acesso ao mercado.

Embora, por definição, o *call center* seja uma forma de relacionamento sem contato presencial, seu uso vem se tornando cada vez mais intenso por empresas de diversos setores da economia. A indústria, o comércio e uma infinidade de empresas de serviços utilizam-se deste tipo de recurso para atender, vender, contatar e entender melhor seus clientes. (WEISSHUHN, 2004)

Segundo Azevedo e Caldas (2002), um *call center* consiste, por essência, em uma atividade do setor de serviços, que funciona como um elo entre a empresa e a sua base de clientes. Os autores acrescentam que as inovações tecnológicas nesse setor têm sido muito grandes, provocando uma reestruturação ampla no funcionamento dessas organizações.

Todo consumidor quer ser bem atendido e o mercado de hoje exige um atendimento cada vez mais ágil e personalizado. E as grandes corporações mundiais já encontraram a resposta para essas demandas: *Contact Center*, que consiste em muito mais do que o *call center* tradicional. Este permite apenas o contato do cliente através de telefone. O *contact center* permite o contato do cliente através de diversos meios, dentre eles o próprio telefone, fax, carta e Internet, com a opção sendo escolhida de acordo com a sua própria preferência, proporcionando o mesmo nível de serviço e de consistência de informações entre eles. Além disso, o *contact center* vem sendo encarado como o principal meio de viabilização da estratégia de CRM das empresas. (CONTAX, 2006)

Chokshi (1999) diz que as empresas estão migrando seus cuidados nas operações com clientes para a Internet, investindo em "cuidado eletrônico", ferramentas de automação de vendas etc.. Uma quantidade significativa dos investimentos em sistemas humanos e de computadores está colocada no *set-up* de *call centers*.

Call centers podem ser do tipo *inbound* (onde as chamadas são originadas pelos clientes), *outbound* (onde as chamadas são originadas pelos operadores para os clientes, geralmente para fins de *telemarketing*), ou misto (onde ambas as atividades são executadas). Para o primeiro tipo, o volume de chamadas e a sua duração caracterizam-se como dados de entrada; para o segundo tipo, consistem em variáveis de decisão. (GROSSMAN *et al*, 2001)

Hawkins *et al* (2001) dizem que *call centers* do tipo *inbound* são caracterizados como um sistema constituído de várias pessoas (atendentes), que recebe ligações telefônicas de outras pessoas (normalmente clientes – mesmo que potenciais) que desejam obter alguma informação, contratar algum serviço, comprar algum produto, responder alguma pesquisa, atualizar dados, registrar ocorrências, ou fazer reclamações, entre outros.

Para eles, as centrais deste tipo são tipicamente vistas como fornecedores de uma gama particular de serviços aos clientes que fazem contato com ela, enquanto as do tipo *outbound* são geralmente caracterizados como centros de lucro (por exemplo, uma central de cobrança).

Os atendentes dos *call centers* de serviço ao cliente tendem a ter diferentes habilidades (básicas e secundárias); alguns podem lidar com um tipo de chamada, enquanto outros podem lidar com outros tipos de chamada. Muitas vezes, pode haver especialização por parte dos agentes. Avanços nos distribuidores automáticos de chamadas (ACDs) tornaram possível o roteamento baseado em habilidades (*skill-based routing*), que consiste no protocolo para o roteamento *on-line* das chamadas entrantes para os agentes apropriados a atendê-las. (MAZZUCHI & WALLACE, 2004; DECKER, 1999; PICHITLAMKEN *et al*, 2003; KOOLE & POT, 2005; KLUNGLE & MALUCHNIK, 1997; TAKAKUWA & OKADA, 2005)

Na economia global atual, pode ser eficiente (em termos de custo) espalhar o trabalho das centrais de serviço ao cliente através de múltiplas cidades, países ou continentes. O obstáculo para atingir essa eficiência reside no dimensionamento de que parcela da carga de trabalho deve ser enviada a cada localidade (incluindo a definição dos critérios nos quais a distribuição das chamadas deve se basear) e quantos agentes seriam requeridos para lidar com tal carga. Complicações adicionais incluem um desejo potencial de compartilhamento e de priorização de determinados centros para lidar com determinadas chamadas. (RILEY, 2005; DECKER, 1999; MILLER & BAPAT, 1999)

O *call center* pode pertencer à empresa prestadora do serviço (ou vendedora do produto) ou ser terceirizado. Existem empresas que oferecem para outras o serviço de atendimento através de uma central. (GROSSMAN *et al*, 2001).

2.1.1. *Call centers* no Brasil

A estabilidade econômica iniciada na década de 90, o crescimento do segmento de serviços e as privatizações das telecomunicações abriram espaço para o desenvolvimento da indústria de *call centers* no país.

Além disso, o consumidor brasileiro vem mudando: tem menos tempo para as compras, gosta de resolver suas pendências pelo telefone e tem muita facilidade para trocar de marca. Por isso, e segundo Minghelli (2002), o mercado de *call centers* aumenta gradativamente a cada ano no Brasil. Lima (2006) diz que as empresas estão cada vez mais dispostas a melhorar o relacionamento com os clientes.

A privatização do setor de telecomunicações recebeu como herança estatal uma memória de maus serviços na percepção do consumidor, que inclui o atendimento ao cliente. O setor tem o desafio de reverter essa insatisfação ao mesmo tempo em que sofre pressões constantes por redução de custos e esbarra nas limitações educacionais do mercado de trabalho e do mercado consumidor, o que dificulta a adoção de novas tecnologias. (GOMES, 2004)

De acordo com especialistas do setor de *call centers*, o ano de 2003 deu início à consolidação dos serviços de tele-atendimento no país. A movimentação financeira oriunda das oportunidades de terceirização no mercado brasileiro de *call center* deverá revelar uma significativa taxa média anual de crescimento até 2008. (BRASIL fecha 2003 com mais de 68 mil Pas, 2005)

Segundo (CRESCE faturamento do setor de telemarketing, 2005), esse mercado cresceu perto de 30% em 2003 – aumentando seu faturamento de R$ 1,96 bilhão para R$ 2,51 bilhões – e em torno de 20% em 2004, faturando aproximadamente R$ 3 bilhões. O vigor é atestado por outro indicador: o número total de posições de atendimento (PAs), que já aumentara 6,2% em 2003 – crescendo de 183,0 mil para 194,3 mil – teve uma evolução ainda mais acentuada em 2004 (crescimento de 12,4%, passando para 218,4 mil PAs).

Uma pesquisa publicada (EMPREGO em contact center está em alta, 2005) recentemente também aponta para um ritmo acelerado de contratações, acompanhando o crescimento do número de posições de atendimento terceirizadas na América Latina, que deve saltar para mais de 200 mil em 2008. Segundo analistas do setor, só o Brasil será responsável por 53,4% desse montante, o que significará a criação de milhares de vagas no mercado.

Dessa forma, o segmento tem merecido especial atenção tanto por seu rápido crescimento, quanto pela grande quantidade de empregos gerados. No ano de 2000, o setor empregava cerca de 370 mil pessoas, tendo este número chegado a mais de 550 mil em 2004. Apenas em 2004 foram gerados 55 mil novos empregos neste setor. A Associação Brasileira de Telemarketing (ABT) diz que o volume financeiro estimado de transações apoiadas por serviços de *telemarketing*, no Brasil, chega a R$ 67,4 bilhões. (WEISSHUHN, 2004)

No início de 2006, o número de operadores atuando no segmento atingiu 650 mil. Baseada no crescimento experimentado no ano anterior, a ABT previu para este ano uma abertura de 50 mil vagas no setor. (LIMA, 2006)

O crescimento acelerado é resultado de dois fatores, segundo especialistas: "O mercado está obrigando as empresas a se reestruturarem para melhorar o relacionamento com o cliente. As companhias perceberam que podem ter o SAC e a venda à distância sob o mesmo investimento.". E as perspectivas para o futuro desse mercado são ainda mais promissoras, já que a demanda vem crescendo e o país já possui uma boa infra-estrutura telefônica disponível e necessária para esse tipo de serviço. (CRESCE faturamento do setor de telemarketing, 2005)

Segundo levantamento recente (BRASIL está entre os 10 países mais atraentes para call centers, 2005), o Brasil está em 7º lugar entre os países mais atraentes para a instalação de centros de atendimento e prestação de serviços de empresas internacionais à distância, no sistema chamado de *business process outsourcing*. O relatório mostra que o país está na mais alta posição entre os países da América Latina em termos de pessoas capacitadas e de disponibilidade. Os pontos fortes do Brasil são os custos vantajosos e uma grande força de trabalho com relativamente boa experiência em prestação de serviços à distância. O estudo ainda aponta a possibilidade de, com a crescente sofisticação da indústria local de *software* e do setor de informação e tecnologia, o país atrair novos negócios.

Minghelli (2002) apresenta os resultados de um estudo que pesquisou quais são os principais segmentos de atuação dos *call centers* brasileiros e as suas estimativas de crescimento (evidenciados na tabela 1 a seguir, onde os campos em branco correspondem a valores que não estavam disponíveis).

Tabela 1 – Magnitude atual e crescimento dos principais segmentos de atuação dos *call centers* no Brasil

Segmento	Posições de atendimento em 2001	Taxa de crescimento histórico	Taxa de crescimento projetado para 2004
Telefonia fixa	19.895	17,6%	8,5%
Cartões	8.738	20,9%	20,8%
Telefonia celular	4.415	35,8%	15,4%
Utilidades	3.490	3,5%	3,3%
Internet	3.000		23,0%
Bancos	2.400		
TV por assinatura	1.492	9,0%	
Saneamento	1.354		
Total	44.784		

Fonte: MINGHELLI (2002)

Como pode ser observado, o ramo de telefonia fixa é o que mais utiliza *call centers*; mas os ramos de cartões, telefonia celular e Internet são os que apresentam maior potencial esperado de crescimento.

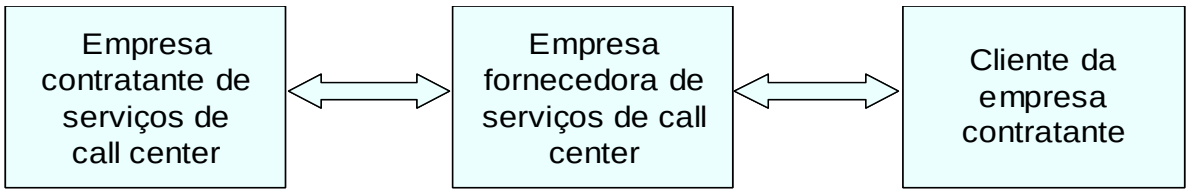
O mercado está concentrado em grandes cidades do país, como São Paulo e Rio de Janeiro. As maiores empresas de *call center* no cenário brasileiro são a Contax, a Atento e a Teleperformance (nessa ordem, em termos de posições de atendimento; e trocando as duas primeiras de lugar, em termos de faturamento e número de funcionários). (OUTSOURCING, 2005)

De acordo com (EMPREGO em contact center está em alta, 2005), o setor emprega principalmente jovens entre 16 e 24 anos, a maioria deles do sexo feminino, em início de carreira. A média salarial varia entre US$ 120 e US$ 200 mensais por seis horas diárias de serviço, podendo alcançar US$ 600 no caso de funcionários bilíngües ou com mais experiência.

Para falar a respeito do segmento de *call centers* no Brasil, é preciso separá-lo em duas partes: *call centers* próprios e *call centers* terceirizados. O primeiro formato entende-se como sendo a estrutura criada e operada por uma empresa para relacionar-se com seus próprios clientes. O segundo formato consiste em uma estrutura criada e operada por uma empresa, especializada neste tipo de serviço, para funcionar como meio de relacionamento entre outras empresas e seus clientes.

Isto quer dizer que as empresas que atuam no mercado de terceirização de *call centers* especializam-se em conhecer e atender os clientes de seus clientes, formando uma estrutura de relacionamento conforme descrita na figura 2 a seguir. (WEISSHUHN, 2004)

Figura 2 - Estrutura de relacionamento entre as entidades participantes na indústria de *Call Centers*



Fonte: WEISSHUHN (2004)

Segundo Mancini (2001) e Freitas (2000), a terceirização do serviço de atendimento telefônico pode ser:

- total – sendo a empresa contratada responsável pelas instalações físicas, *softwares*, sistemas de telefonia, computadores e mão-de-obra;
- parcial – sendo a empresa contratada responsável pela mão-de-obra, trabalhando nas instalações da empresa contratante, ou vice-versa; ou
- mista – sendo a infra-estrutura pertencente à empresa contratada e a mão-de-obra constituída de uma mistura de supervisores e coordenadores da empresa contratante com operadores da empresa contratada.

A terceirização é uma alternativa a ser considerada na decisão de implantação de um *call center*, cabendo avaliar a relação de custo-benefício de terceirizar ou manter o *call center* como um setor da empresa. (MINGHELLI, 2002)

De acordo com Mancini (2001), não existe um sistema absolutamente melhor do que o outro: cada caso é um caso; cada empresa atravessa um momento em seu ciclo de vida, exigindo respostas personalizadas. Os principais benefícios da terceirização são: redução de tempo; readaptações estratégicas mais rápidas; maior controle através de relatórios consolidados; e especialização.

A predominância de *call centers* próprios é habitualmente atribuída ao fato de a maioria das empresas considerar este canal de relacionamento com o cliente como uma função estratégica, onde a terceirização pode representar um risco à obtenção dos objetivos para a qual a operação foi concebida, já que o grau de compromisso de funcionários de uma empresa contratada pode não ser o mesmo que se esperaria de funcionários próprios. Muitas empresas têm dúvidas quanto à contratação de empresas para realizar o atendimento ao seu cliente, pois a área é sensível e envolve, na maior parte das vezes, necessidade de sigilo.

Conforme Mancini (2001), quando ocorre a contratação de uma empresa terceirizada para realizar os serviços de *call center*, a empresa contratada deverá manter uma área específica e exclusiva para executar as ações da empresa contratante, mantendo assim a qualidade e a confidencialidade necessárias para a operação.

Contudo, algumas pesquisas vêm evidenciando mudanças nessa predominância de centrais próprias, o que pode ser notado pelos resultados que indicam maior crescimento percentual do número de posições de atendimento em *call centers* terceirizados do que em *call centers* próprios.

O serviço oferecido pelas empresas de *call center* engloba uma ampla gama de soluções em áreas como SAC (serviço de atendimento ao cliente), tele-cobrança, *telemarketing* ativo e receptivo, retenção e fidelização de clientes, entre outros. A diversidade de serviços e, principalmente, a customização oferecida pelas empresas que disputam este mercado têm sido apontadas como os principais fatores na conquista de novos contratos, o que tem levado o mercado a apostar em um duradouro ciclo de crescimento para este setor. (WEISSHUHN, 2004)

Em função da relevância percebida por grande parte das empresas em relação aos serviços de *call center* como um canal estratégico de relacionamento com seus clientes, aquelas que se apresentam dispostas a terceirizar tais operações tratam este processo como uma aquisição

estratégica de serviços, procurando estabelecer contratos de parceria, onde o conhecimento recíproco entre empresa contratante e empresa fornecedora de serviços consiste em um aspecto fundamental.

Segundo (CONTAX, 2006), apesar dos avanços, os serviços de atendimento ao cliente continuam em transformação, embora já sejam vistos como parte fundamental do processo de conquista e de fidelização dos consumidores. A evolução no atendimento pode ser dividida em quatro eras:

1. Atender: implantação e planejamento básico;
2. Entender: conhecer melhor cada tipo de consumidor e entender suas demandas e necessidades;
3. Resolver: redesenhar os processos internos para atender – no primeiro contato – as necessidades do cliente;
4. Antecipar e vender: pró-atividade que visa antecipar-se às necessidades do cliente e aproveitar cada contato como oportunidade de oferecer novos produtos e serviços.

O desejo das empresas é estar nesta última fase. Com todas as mudanças ocorridas nos últimos anos, as companhias se depararam com os desafios de gerenciar o crescimento de seus *call centers* e transpor as eras.

A ANATEL tem, entre os seus objetivos, o de estimular a competição e a melhoria do serviço prestado ao consumidor do setor de telecomunicações. Ela começou a atuar após a privatização e seu conjunto de indicadores e metas mudou pouco desde então. No início, os indicadores consistiam em metas até agressivas para empresas recém-privatizadas e com grande potencial de evolução. Atualmente, a maioria dos indicadores foi atendida pelas principais empresas do setor e a grande pressão por melhoria de desempenho vem da concorrência e dos clientes, e não da regulação do governo.

O preço é, normalmente, o principal aspecto que irá conduzir a escolha da empresa que irá prestar o serviço de *call center*. Ele pode ser visível para a empresa contratante através de três formas diferentes de faturamento:

- *speaking time* – faturamento baseado nos minutos falados pelos atendentes. O cálculo é feito multiplicando o volume de chamadas pelo Tempo Médio de Atendimento (TMA) e pelo preço do minuto falado;

- por posição de atendimento (PA) – faturamento baseado na quantidade média de PAs ocupadas nos dias úteis no mês, que pode apresentar três modalidades:

    - PA Completa – Fornecimento da mão-de-obra, infra-estrutura, etc.;

    - PA Mão-de-obra – Fornecimento apenas dos operadores;

    - PA Infra-estrutura – Fornecimento apenas do espaço.

- por resultado – faturamento baseado na efetividade do trabalho dos atendentes (quantidade de dívidas recuperadas por uma central de cobranças, por exemplo).

O primeiro formato é mais utilizado pelos maiores clientes; as menores empresas geralmente preferem o segundo formato, que é o mais comum entre todos. O terceiro formato é encontrado, normalmente, em *call centers* do tipo *outbound*.

Geralmente, a empresa contratante exige um nível mínimo de serviço da empresa contratada através do *Service Level Agreement* (SLA). Metas (90% das ligações devem ser atendidas em até 10 segundos, por exemplo) são estabelecidas em conjunto e a empresa contratada é penalizada no seu faturamento em caso de falhas no seu cumprimento.

Um fator de extrema relevância costuma surgir no funcionamento dos processos de formação de preços: o acirramento da competição entre aquelas empresas que dominam atualmente este mercado. Isto tem levado as empresas a comprimirem suas margens de lucro até o limite da rentabilidade mínima aceita pelos acionistas para este tipo de negócio. Este fenômeno se torna especialmente crítico quando ocorre a disputa por clientes considerados estratégicos. Exemplos disso podem ser observados nos processos de terceirização dos *call centers* de companhias pertencentes a grandes grupos empresariais atuantes no país, situação onde os concorrentes tendem a dar importância especial à conquista do cliente, por considerá-lo uma porta para alcançar as demais empresas do mesmo grupo. (GOMES, 2004)

Esse acirramento da competição é ainda mais freqüente porque, de acordo com Weisshuhn (2004), o mercado de terceirização de *call centers* apresenta um nível de concentração

relativamente alto. As 5 maiores empresas deste segmento concentram cerca de 60% do número total de posições de atendimento terceirizadas existentes no Brasil e mais de 90% do mercado total pertence às 12 maiores empresas da indústria.

Esta grande concentração de mercado pode ser atribuída à ocorrência de vantagens competitivas relacionadas à associação de algumas empresas de *call center* a grandes empresas prestadoras de serviços de telecomunicações, pois, ao mesmo tempo em que estas associações representam acesso a tecnologias e serviços de comunicação de dados e voz a custos competitivos, também são elas próprias as maiores clientes do mercado de *call centers*. De fato, as maiores fornecedoras de serviços de *call center* no Brasil são empresas controladas por grandes grupos atuantes na área de serviços de telecomunicações. Assim sendo, estas empresas já têm asseguradas como clientes as demais empresas do mesmo grupo. Isto lhes permite obter escala e conhecimento técnico para atuação no segmento independentemente da conquista de outros clientes, além de lhes garantir, por causa do porte de suas controladoras, os recursos financeiros para a conquista e o atendimento de contratos de grande magnitude, que exigem investimentos mais pesados em infra-estrutura e equipamentos de comunicação de dados e de voz. (GOMES, 2004)

Essa é exatamente a situação em que se encontra a empresa estudada nesta pesquisa: a Contax, pertencente ao grupo Telemar, e que será mais bem descrita na seção 2.2 a seguir.

2.2. A Contax

2.2.1. Histórico da empresa

A Contax surgiu no final de 2000, como extensão natural dos negócios da Telemar – em um setor da economia que pouco investia em tecnologia e qualificação de atendimento – para ajudar seus clientes na gestão operacional do serviço de atendimento, agregando valor no relacionamento com os consumidores finais. (CONTAX, 2006)

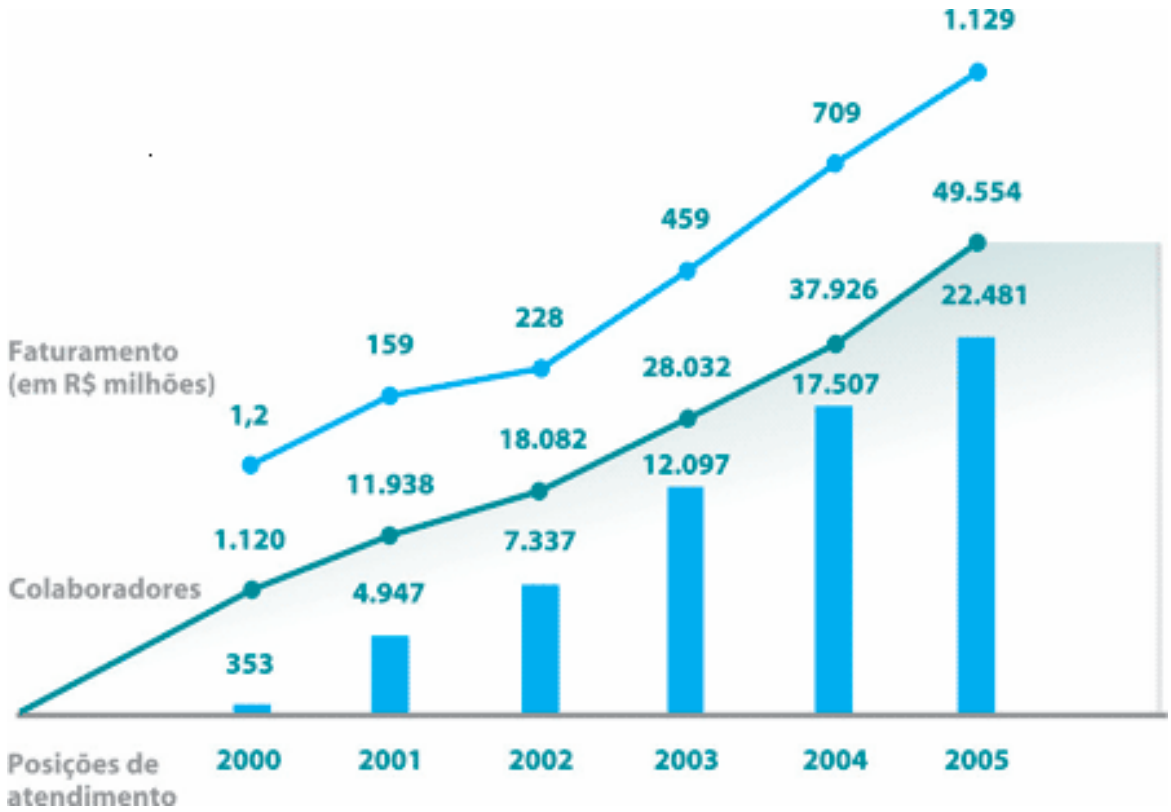
A companhia rapidamente tornou-se expoente desse mercado, oferecendo diferenciais que possibilitaram atrair clientes que buscavam alavancar a venda de seus produtos e agregar valor no relacionamento com seus consumidores. Para alcançar esse nível de excelência operacional, a Contax direcionou seu trabalho para a produtividade e para os serviços personalizados que projeta, implementa e opera. Foram esses conceitos que permitiram que a

empresa ganhasse respeito no segmento. A companhia conseguiu, em apenas 4 anos, o posto de uma das maiores geradoras de emprego do país e aumentar seu faturamento bruto em quase 300% .

Tamanho sucesso não seria sustentável caso a Contax deixasse de lado a capacitação pessoal de seus funcionários e a tecnológica. Mais de R$ 250 milhões foram aportados para essas áreas e, assim, foi possível equipar a empresa com modernas centrais, munidas com as melhores e mais adequadas ferramentas tecnológicas e estruturas de *contact center*. Isso também permitiu preparar profissionais altamente capacitados, que garantem um atendimento que satisfaça todos os públicos envolvidos no negócio. (CONTAX, 2006)

A Contax tem se mantido em constante crescimento desde a sua criação em 2000, o que pode ser evidenciado no gráfico apresentado na figura 3 a seguir.

Figura 3 - Evolução do faturamento, do número de colaboradores e de posições de atendimento da Contax, 2000 a 2005



Fonte: CONTAX (2006)

Segundo a mesma fonte, os principais motivos que permitiram a manutenção deste crescimento foram:

- foco em operações de grande volume de transações e alta complexidade;
- foco em qualidade, com o objetivo na satisfação do cliente final (investimento massivo em treinamento e processos);
- capacidade de investimento.

2.2.2. Números atuais da empresa

A Contax é uma empresa do grupo Telemar, atuando no mercado de *contact centers*. No Brasil, atualmente, é a empresa de maior crescimento nessa indústria, tendo crescido quase 60% em 2005 (com um faturamento de R$ 1.129 milhões). Segundo (OUTSOURCING, 2005), é a maior empresa do ramo em número de posições de atendimento, e a segunda maior em termos de faturamento e número de funcionários, dentro do território nacional.

De acordo com (CONTAX, 2006), ela faz mais de 100 milhões de contatos por mês, através de telefone, correio, *e-mail*, torpedos, Internet e *chat*, e conta com mais de 22 mil posições de atendimento e quase 50 mil funcionários distribuídos em 16 centrais por 7 estados brasileiros, além do Distrito Federal: Bahia, Pernambuco, Ceará, Rio Grande do Sul, Minas Gerais, Rio de Janeiro, São Paulo (compondo o mais moderno complexo de *contact centers* da América Latina, na capital). Ela está presente nas capitais, e em Niterói (RJ) e Nova Lima (MG).

2.2.3. Produtos e clientes da empresa

Esta abrangência geográfica permite um atendimento regionalizado, que garante a entrega de soluções personalizadas para seus clientes. Com este foco na customização, a Contax desenvolveu e implementou um abrangente leque de soluções e serviços completos de atendimento, que vai desde o primeiro contato até a integração com sistemas de logística das empresas, incluindo: SAC, *help desk*, *telemarketing* ativo e receptivo, pesquisa de mercado, retenção, *contact center* para contingências, serviços de cobrança e Internet *call center*. Assim, mais do que uma prestadora de serviços de *contact center*, a Contax atua como consultora (e co-gestora) de seus clientes para identificar a melhor solução de relacionamento com os seus próprios consumidores. (CONTAX, 2006)

A diversidade dos clientes está presente também no quesito forma de faturamento. Eles podem ser cobrados de três maneiras distintas, a critério do próprio cliente:

- *speaking time* – faturamento baseado nos minutos falados pelos atendentes: volume de chamadas multiplicado pelo TMA e pelo preço do minuto falado;

- por posição de atendimento (PA) – faturamento baseado na quantidade média de PAs (que podem ser completas, incluir apenas a mão-de-obra ou apenas a infra-estrutura) ocupadas no horário mais movimentado dos dias úteis do mês;

- por resultado – faturamento baseado na efetividade do trabalho dos atendentes (quantidade de dívidas recuperadas por um *call center* de cobranças, por exemplo).

O formato mais comum na empresa é o segundo, preferido pelos menores clientes da Contax. O primeiro formato é o mais utilizado pelos maiores clientes. O terceiro formato é aplicado, normalmente, a clientes que contratam a operação de um *call center* do tipo *outbound*.

Os clientes costumam exigir – através do *Service Level Agreement* (SLA) – um percentual mínimo de ligações a serem atendidas dentro de uma quantidade pré-determinada de tempo. Estas metas são estabelecidas em conjunto (por exemplo: 95% das ligações devem ser atendidas em até 15 segundos) e a Contax é penalizada no seu faturamento em caso de falhas no seu cumprimento. Vale destacar que este percentual é conhecido internamente como "nível de serviço". No caso do formato de faturamento por *speaking time*, é comum o tempo médio de atendimento fazer parte do SLA (não podendo ser superior a 3 minutos, por exemplo).

A Contax é uma empresa de capital 100% nacional e hoje opera com mais de 40 clientes (a maioria líder de mercado ou com atuação destacada em seu segmento), como: Nossa Caixa, Grupo Unibanco, Grupo Banco do Brasil, Citibank, Bradesco, HSBC, Caixa Econômica Federal, ASB Financeira, Hipercard, Losango, Itaucard, GE, Banco Real ABN AMRO, Finasa, Sul América Capitalização, Xerox, Cemig, NET, Ticket, Coelba, TecBan, Pernambucanas, UOL, Terra, Correios, Oi, IG, entre outros. (CONTAX, 2006)

Seu principal cliente é a própria Telemar (que responde por aproximadamente 60% do faturamento), sendo que os principais produtos estão relacionados e descritos a seguir:

- 102 – que recebe ligações dos clientes que buscam auxílio à lista telefônica;

- Velox – que engloba o suporte técnico e o *help desk* para os clientes do serviço de Internet de banda larga da Telemar;

- 103 – que abrange clientes que buscam serviços relacionados à telefonia fixa, como mudança de endereço de cobrança e dúvidas ou problemas com a conta telefônica;

- Suporte técnico e reparo de defeitos da Telemar – cuja entrada do cliente acontece via 103 (o que é melhor para ele, que disca um único número para qualquer tipo de atendimento), sendo este desviado para a central apropriada através da URA de voz, não causando, assim, impacto nem congestionamento das filas do produto 103;

- OI – que abrange todo o atendimento para os clientes dos serviços de telefonia móvel da Telemar.

A companhia tem como missão ser a primeira escolha para construir relacionamentos entre empresas e pessoas, visando fazer com que o *contact center* passe de centro de custos para centro de lucros de seus clientes. Este é o principal motivo para a empresa já ter recebido diversos prêmios em relação a aspectos como atendimento, comunicação, tecnologia e qualidade.

Esta busca pela satisfação do cliente, no entanto, ainda é realizada de forma ineficiente, na opinião da equipe de operação da empresa, que admite que os custos para atingir tal *status* frente aos clientes ainda estão altos, apresentando grande potencial de redução. Tal constatação posiciona a empresa no quadrante superior esquerdo (efetivo, mas ineficiente) da matriz de dimensões no atendimento ao cliente de *call center* (apresentado anteriormente na figura 1), clamando por melhorias operacionais capazes de fazer a posição da empresa mudar para o quadrante superior direito da mesma figura (efetivo e eficiente).

3. Referencial teórico

3.1. Métodos de previsão de demanda

Uma grande ênfase tem sido dada à melhoria do processo de tomada de decisão em organizações governamentais e empresariais. Por causa do ambiente fortemente competitivo, tornou-se objetivo básico alocar de maneira otimizada o pouco tempo disponível entre atividades concorrentes. Esse objetivo pode ser alcançado, em parte, por meio da previsão precisa. Nos anos 70, os administradores tomavam suas decisões baseando-se fortemente em seus próprios sentimentos e intuições sobre a indústria e a economia. Hoje eles estão complementando esse "sentimento" em relação à indústria ou à economia com a utilização de técnicas de previsão simples e sofisticadas. A difusão do uso de microcomputadores continua a reforçar essa tendência. (HANKE, 1992)

Em um ambiente organizacional em contínua transformação, o que se procura é reduzir incertezas. Os gestores demandam informações que os auxiliem a escolher, hoje, as que parecem ser as melhores alternativas sobre eventos que ocorrerão no futuro. Permitir a antecipação de cenários futuros é a proposta dos modelos de previsão. (CORRAR & THEÓPHILO, 2004)

Uma vez que as condições econômicas e de negócios variam ao longo do tempo, os gerentes de negócios precisam encontrar maneiras de se manter a par dos efeitos que essas mudanças terão em suas operações. Uma técnica que eles podem empregar, como forma de ajuda ao planejamento de necessidades operacionais futuras, é a previsão. Embora seja grande o número de métodos de previsão disponíveis, todos têm um objetivo em comum – fazer previsões sobre eventos futuros, de modo que essas projeções possam ser incorporadas ao processo de tomada de decisão. (LEVINE, BERENSON & STEPHAN, 2000)

Levine, Berenson e Stephan (2000) dizem que existem, basicamente, dois tipos de métodos de previsão: o qualitativo e o quantitativo. O primeiro tipo é especialmente importante quando dados históricos não estão disponíveis, mas é visto como altamente subjetivo e passível de avaliação.

Já os métodos quantitativos podem ser subdivididos em dois tipos: de série temporal e de série causal. O segundo tipo – dentro do qual a regressão múltipla consiste em um exemplo

clássico – envolve a determinação de fatores que se relacionam com a variável a ser prevista. Os métodos de previsão de uma série temporal – que é um conjunto de dados numéricos obtidos durante períodos regulares ao longo do tempo – envolve a projeção de valores futuros de uma variável, com base, inteiramente, em observações do presente e do passado dessa variável.

Em qualquer dos tipos, um interessante indicador capaz de mensurar a magnitude do erro de previsão incorrido pelo método é o **E**rro **M**édio **P**ercentual **A**bsoluto (MAPE). Ele consiste na média dos módulos dos erros percentuais de previsão, sendo estes últimos obtidos para cada valor previsto através da fórmula EP (erro percentual) = (valor previsto – valor real) / (valor real). (HANKE, 1992)

3.1.1. Séries temporais

A principal característica de uma série temporal, que a torna distinta de uma amostra aleatória simples, é a vinculação das observações com o fator tempo. (CORRAR & THEÓPHILO, 2004)

Silver (2000) justifica a importância dos métodos de previsão para séries temporais dizendo que, em muitas análises feitas nas ciências econômicas e administrativas, existe interesse particular em saber como uma variável se comporta ao longo de tempo.

O pressuposto básico da análise dessas séries, segundo Levine, Berenson e Stephan (2000), é que os fatores que influenciaram padrões da atividade no passado e no presente continuarão a fazê-lo, mais ou menos da mesma maneira, no futuro.

Por isso, Smailes e McGrane (2002) alertam que se deve tomar cuidado ao prever o futuro com estes modelos, já que o que está sendo realizado é uma extrapolação.

Hanke (1992) sugere uma estratégia típica para avaliar os métodos de previsão, envolvendo as seguintes etapas:

1. Um método de previsão é escolhido baseando-se na intuição do analista sobre o padrão dos dados.

2. O conjunto de dados é dividido em duas seções – uma parte para inicialização e outra para teste.

3. A técnica de previsão escolhida é usada para desenvolver parâmetros ajustados à parte de inicialização dos dados.

4. A técnica é usada para prever a parte de teste dos dados, sendo o erro de previsão calculado e avaliado.

5. Uma decisão é tomada: empregar a técnica em sua forma atual, modificá-la, ou calcular uma nova previsão usando outra técnica (a comparação deve ser sempre baseada nos resultados alcançados, em termos do erro de previsão).

Já Levine, Berenson e Stephan (2000) oferecem três métodos como linhas mestras para a seleção do modelo a ser utilizado:

1. Fazer uma análise de resíduos, que consistem na diferença entre os dados observados e os dados ajustados. Se o modelo em questão se ajustar adequadamente, os resíduos representarão o componente irregular da série temporal e devem, portanto, ser distribuídos aleatoriamente ao longo da série; se ele não se ajustar adequadamente, os resíduos poderão demonstrar um padrão sistemático, deixando de levar em conta, por exemplo, uma tendência ou variação sazonal;

2. Mensurar a magnitude do erro residual. Se, depois da análise de resíduos, dois ou mais modelos parecerem adequados, a média aritmética dos desvios (ou resíduos; ou ainda erros de previsão) absolutos deve ser calculada para os modelos candidatos e aquele com a menor dessas médias pode ser selecionado como o mais apropriado;

3. Utilizar o princípio da parcimônia. Se, depois dos dois passos anteriores, dois ou mais modelos parecerem adequados, deve ser selecionado o modelo mais simples que execute a tarefa satisfatoriamente.

Silver (2000) também depõe a favor da parcimônia: "Com relação aos modelos (...), sabe-se que os mais simples (...) são tão bons como os mais complexos (...) e têm a vantagem de serem mais fáceis de entender; isso torna provável que os gerentes possam ajustar os resultados para fatores não incluídos no modelo, porque eles sabem o que estão fazendo."

O autor afirma que, como princípio, para que o uso seja adequado, é crítico o conhecimento das pressuposições implícitas nos modelos. Pode então ser usado o julgamento, para ajustar o modelo aos fatores não incluídos e para avaliar o efeito de quaisquer pressuposições que o

método, aparentemente, não atinja. Os modelos simples podem ser muito eficientes, desde que incorporem, adequadamente, o que deve ser medido e deixem o usuário ajustar a precisão, de acordo com o próprio julgamento. Na escolha dos métodos, devem ser considerados: a finalidade da previsão, até quando ela será feita, a relação número-custo das previsões necessárias, o conhecimento dos analistas e usuários, a periodicidade, a compatibilidade com o sistema de planejamento existente, as propriedades estatísticas das séries e o número de observações passadas.

Levine, Berenson e Stephan (2000) falam da importância do monitoramento, de modo apropriado, do modelo de previsão, uma vez que ele for selecionado; afinal, o objetivo do processo de seleção é escolher um modelo que seja capaz de projetar ou prever futuros movimentos em um conjunto de dados da série temporal. Mas, infelizmente, tais modelos de previsão geralmente são precários no que diz respeito a detectar alterações na estrutura básica da série temporal. É importante que essas projeções sejam examinadas e, tão logo um novo dado seja observado em algum período futuro de tempo, ele deve ser comparado com a sua projeção. Se a diferença for muito grande, o modelo de previsão deve ser revisado.

Os métodos de previsão de séries temporais mais utilizados na literatura são:

1. Amortecimento exponencial simples
2. Método de Holt (amortecimento exponencial ajustado pela tendência)
3. Modelo de Winter (amortecimento exponencial ajustado pela tendência e pela variação sazonal)
4. Regressão linear simples
5. Regressão quadrática
6. Regressão exponencial

Hanke (1992) e Silver (2000) fornecem uma descrição aprofundada do primeiro método da lista. A primeira destas referências também apresenta detalhadamente os métodos 2 e 3, enquanto Levine, Berenson e Stephan (2000) visitam a fundo os 3 últimos métodos da lista.

3.1.2. Séries causais

O método mais recorrente na literatura, dentre os integrantes da família dos métodos de série causal, é o da regressão linear múltipla, que – pelo fato de ter sido utilizado para a resolução de um dos principais problemas levantados na empresa estudada – será apresentado a seguir.

3.1.2.1. Regressão linear múltipla

Em geral, se a variável a ser predita está fortemente correlacionada com outras variáveis que estão sujeitas a grandes variações, um modelo multivariado é necessário. (SILVER, 2000)

Levine, Berenson e Stephan (2000) colocam que um modelo de regressão linear múltipla permite prever o comportamento de uma variável (dependente), $\hat{Y}$, em função do comportamento de *n* outras (independentes), X_1, X_2, ... , X_n, através da equação:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + ... + b_n X_{ni},$$

onde b_0 é a interseção, b_1 é a taxa de crescimento da variável Y por unidade da variável X_1, ... , b_n é a taxa de crescimento da variável Y por unidade da variável X_n.

Estes coeficientes podem ser obtidos facilmente em programas de computador, como o Excel, através do método dos mínimos quadrados. Para prever o valor da variável dependente Y em função dos valores das variáveis explicativas X_1, X_2, ... , X_n, basta resolver a equação acima. (LEVINE, BERENSON & STEPHAN, 2000)

Segundo Lapponi (2005), o coeficiente de determinação ajustado (ou R-quadrado ajustado) representa a medida de qualidade para os modelos de regressão linear múltipla. Enquanto o coeficiente de determinação (R-quadrado) consiste no percentual de variação explicado pelo modelo, o seu valor ajustado tenta compensar o aumento natural de explicação provocado pelo aumento do número de variáveis independentes – que ocorre no caso da regressão múltipla – e do tamanho da amostra.

Ainda de acordo com o autor, outro indicador muito importante é o valor F, que testa a hipótese de que nenhum dos coeficientes da regressão tenha significado. Adicionalmente, o valor F de significação consiste no *p-value* do teste; em outras palavras, na probabilidade de ocorrer um alinhamento linear tão bom quanto o que efetivamente ocorreu, sem que as variáveis independentes estejam efetivamente influenciando o comportamento da variável

dependente. Assim, se este valor for menor que o nível de significância desejado, haverá evidências de que a regressão deverá ser aceita.

Individualmente, cada variável independente também pode (e deve) ser submetida a um teste de hipóteses para verificar a sua representatividade. A hipótese a ser testada é a de que o coeficiente angular da variável independente em questão é zero; ou seja, de que a variável não funciona como bom explicador linear para a variável dependente (já que esta não estaria dependendo da suposta variável explicativa). O *p-value* de cada variável traduz a probabilidade do coeficiente em questão assumir o valor nulo e hipótese de não representatividade de cada variável independente deve ser rejeitada quando o seu *p-value* for baixo, ou menor do que o nível de significância estabelecido para o teste.

De acordo com Silver (2000), pode-se incorporar sazonalidade (ou outros fenômenos que alterem o padrão temporal de uma série, como feriados, promoções etc.) ao modelo de regressão por meio de "variáveis binárias".

Também conhecidas como "variáveis simbólicas" ou "variáveis *dummy*" segundo Levine, Berenson e Stephan (2000), elas são úteis quando não se verifica o pressuposto das variáveis explicativas (independentes) serem numéricas. Existem muitas situações em que variáveis categorizadas precisam ser incluídas no processo de desenvolvimento do modelo. Por exemplo, o fato de um dia consistir ou não em um feriado poderia ser representado por uma variável simbólica, Fer, que utilizaria os valores 0 e 1 para indicar as duas categorias às quais um dia poderia pertencer: comum ou feriado, respectivamente. A variável Fer seria tratada como qualquer outra na regressão múltipla, podendo, no entanto, assumir apenas esses dois valores (0 ou 1). O valor do coeficiente encontrado para essa variável indicaria o efeito incremental (na variável dependente) resultante do fato de o dia ser um feriado.

3.2. Previsão de demanda de *call centers*

Os supervisores dos *call centers* precisam dimensionar eficientemente as suas equipes para fornecer satisfatórios níveis de serviço a custos razoáveis. A administração apropriada das centrais requer a estimação de algumas premissas operacionais capazes de determinar níveis apropriados de tamanho da equipe; entre estas premissas, destaca-se a demanda prevista de ligações de clientes, questão que deve ser abordada regularmente. (WEINBERG, BROWN e STROUD, 2006; KLUNGLE & MALUCHNIK, 1997)

De acordo com Shen e Huang (2005b), Gans, Koole e Mandelbaum (2003) e Klungle e Maluchnik (1997), o gerenciamento efetivo e eficiente da operação de um *call center* – mantendo um nível de serviço consistente – requer um encaixe entre os recursos internos e a carga de trabalho. Para prever esta última de forma acurada, o primeiro – e mais crítico – passo é conseguir uma previsão acurada dos volumes futuros de ligações (ainda mais importante nas centrais com altos níveis de utilização). Usualmente, dois tipos de previsão são necessários para o gerente de um *call center* conseguir dimensionar e planejar os horários da sua equipe:

- Prever os volumes de ligações alguns dias e semanas à frente;
- Atualizar dinamicamente a previsão, em um dia particular, usando novas informações disponíveis à medida que chamadas adicionais chegam à central ao longo do dia. Isto é útil porque gera flexibilidade para a alocação dos recursos disponíveis, acarretando uma maior eficiência e produtividade. Ao reavaliar a previsão para o restante do dia, o gerente pode marcar reuniões ou sessões de treinamento para os atendentes que ficaram repentinamente sem trabalho, ou chamar para o trabalho agentes de *backup*.

Matan e Nourbakhsh (1998) dizem que, para criar previsões acuradas dos volumes futuros de chamadas, os gerentes dos *call centers* precisam juntar informações de várias fontes, desde previsões do mercado até a duração dos intervalos comerciais. No entanto, parte das informações mais importantes vem das estatísticas de volume histórico de chamadas que, usualmente, podem ser ajustadas para as tendências atuais ou esperadas.

Os autores alertam para a necessidade de essas estatísticas serem acuradas. Um erro comum é usar a grandeza "chamadas atendidas" como medida do volume de ligações passadas; mas esta grandeza consiste apenas das chamadas atendidas com sucesso pelos agentes, ignorando as chamadas abandonadas. O correto seria considerar o total de chamadas realizadas, caracterizando a demanda de ligações.

Os modelos de séries temporais têm se constituído no método mais popular para o tratamento da previsão de chamadas telefônicas, segundo Mehrotra (1997). As chamadas são previstas, baseando-se no histórico de dados, mas levando também em consideração o dia da semana e os fatores relativos ao mês do ano. Além disso, outros formatos de modelos de previsão foram aplicados em alguns ambientes de *call center*, com sucesso, como, por exemplo:

- centrais de vendas baseando sua previsão de chamadas nas remessas de produtos recebidas e nas propagandas para os clientes;

- centrais responsáveis pelo suporte de *softwares* baseando sua previsão no número de vendas e *upgrades* de produtos;

- centros de reservas usando modelos de previsão que refletem o impacto de uma guerra de preços.

Segundo Shen e Huang (2005b), abordagens quantitativas para previsão de chegada de chamadas fazem uso dos dados históricos de chegada, que armazenam os momentos no tempo em que ligações chegaram à central. Normalmente, estes dados são agregados em números totais de chegadas durante pequenos intervalos de tempo, como 15 ou 30 minutos; subseqüentemente, o alvo da previsão consiste nos volumes de chamadas atingindo o *call center* durante tais períodos.

A atividade de previsão é inerentemente difícil no caso dos *call centers* por causa do pequeno tamanho dos "blocos de tempo". Técnicas de previsão estabelecidas (como o amortecimento exponencial de Winters e a regressão) são úteis para este tipo de negócio. Geralmente, usar uma abordagem de previsão apropriada irá cortar pela metade os erros de previsão. No entanto, muitos *call centers* encontram dificuldades nesse processo por conta do conhecimento técnico necessário para adequar estas técnicas aos complexos padrões de chamadas e ao aspecto "desarrumado" dos dados. É necessário lidar com picos de demanda (e a perda de clientes oriunda deste fenômeno – tanto por abandono quanto por insucesso no ato de completar a chamada) e lacunas de demanda, que podem ocorrer randomicamente e/ou com magnitudes diferentes a cada momento. (GROSSMAN *et al*, 2001).

Betts, Meadows e Walley (2000) acrescentam que os *call centers* freqüentemente experimentam grandes flutuações na demanda ao longo de períodos de tempo relativamente curtos. Adicionalmente, muitas centrais também precisam manter baixos os tempos de resposta à demanda. Apesar de existir uma forte associação entre as medidas de performance e a abordagem de previsão, nenhum método aparenta ser consistentemente mais efetivo. Os autores deixam clara a importância da acurácia na previsão dizendo que esta tem significativa influência sobre a performance da central, incluindo o grau de sucesso no atendimento das chamadas.

O propósito da previsão é estimar a carga de trabalho de forma a que se possa obter o tamanho certo da equipe para lidar com ela. E existem muitas situações no ambiente de *call center* que requerem uma previsão. O cenário mais comum consiste nas operações normais do dia-a-dia. Mas uma previsão pode ser necessária para situações especiais, como: o planejamento de novos tipos de chamadas; a abertura de um novo centro; uma fusão ou aquisição; uma mudança no horário de operação; a implementação de uma nova tecnologia capaz de afetar o volume ou o padrão das chamadas. Qual for a razão, é importante entender os princípios básicos da previsão da carga de trabalho e como aplicá-los de forma a planejar acuradamente os recursos da central de atendimento. (REYNOLDS, 2005)

A previsão do volume de chamadas é um importante aspecto das operações em *call centers*, sendo necessária para previsão de carga de trabalho, alocação de mão-de-obra e planejamento de capacidade. Existem algumas regularidades dentro dos dados de chegada das chamadas, como a dependência entre os dias (correlação entre os volumes de dias consecutivos e possíveis ciclos semanais, mensais, sazonais e até anuais nos volumes de ligações) e a dependência intradiária (correlação entre chegadas em diferentes períodos de tempo – manhã, tarde ou noite – dentro do mesmo dia). (SHEN & HUANG, 2005a)

Na opinião dos autores, bons modelos de previsão devem capturar pelo menos uma das duas estruturas de dependência ou até mesmo as duas, conforme alguns exemplos existentes na literatura, incluindo o trabalho em questão, dos próprios.

Os mesmo autores contam, em outro artigo (2005b), que os primeiros trabalhos que previam chegadas de chamadas usualmente ignoravam essa dependência intradiária, devido – em parte – à carência de informações relevantes. A partir da virada do século, com a disponibilidade de informações chamada-por-chamada e, impulsionados por uma crescente necessidade de um gerenciamento mais eficiente de *call centers*, alguns modelos estocásticos de chegadas de chamadas que levam em conta a dependência intradiária foram introduzidos. Métodos bayesianos e outros que consideram o efeito do dia da semana, assim como a atualização da previsão intradiária, também apareceram na literatura.

Wise (2005) lembra que o propósito real do processo de previsão é predizer quantos agentes são necessários para lidar com o volume de chamadas. Para produzir uma previsão acurada, os sistemas de gerenciamento da força de trabalho têm que considerar tanto o que aconteceu no passado quanto o que é esperado que aconteça no futuro.

Segundo a autora, todos estes sistemas podem produzir previsões básicas; o problema é que a maioria dos *contact centers* atuais precisam mais do que a funcionalidade básica para prever acuradamente a demanda de ligações para múltiplas centrais, estratégias de roteamento complexas, agentes com múltiplas habilidades e contatos multimídia. Os sistemas de gerenciamento que não conseguem levar em consideração as nuances dos ambientes específicos não são capazes de fornecer previsões acuradas, embaçando a visibilidade das necessidades de agentes e diminuindo a efetividade do negócio.

A mesma, então, relaciona as diferenças entre a funcionalidade de previsão de um sistema básico e a de um sistema mais avançado:

- sistemas básicos oferecem funcionalidade para planejamento apenas do volume de chamadas recebidas, enquanto os sistemas avançados fornecem funcionalidade para planejar contatos não abandonados, como *e-mail* e requisições pela Internet;

- sistemas básicos de previsão utilizam um único tempo médio de atendimento para o dia inteiro, enquanto os sistemas avançados entendem que os valores reais do tempo de atendimento afetam os cálculos de necessidades tanto quanto o volume de contatos e consideram os valores dessa grandeza a cada intervalo de tempo;

- sistemas básicos só conseguem prever ambientes com múltiplas centrais tratando-as como uma grande entidade ou como entidades distintas, enquanto os sistemas avançados são capazes de fornecer o melhor dos dois mundos: prever todas as localidades em conjunto para levar em consideração ganhos de escala, enquanto consideram tempos de atendimento, horas de operação e fusos horários próprios de cada central, acarretando uma visibilidade para as previsões tanto do ponto de vista global como local;

- sistemas básicos têm desafios significativos para prever acuradamente a demanda em ambientes com múltiplas habilidades, enquanto sistemas avançados entendem que, para determinar necessidades em tais ambientes, é necessário considerar durante que percentual de tempo cada agente pode estar disponível para os contatos com os quais ele está habilitado a lidar.

O processo de previsão consiste tanto em arte quanto em ciência. Em arte porque o futuro, afinal de contas, está sendo previsto; e a acurácia da previsão vai depender em parte do

julgamento e da experiência de quem estiver conduzindo o processo. Mas também em ciência: um processo matemático passo-a-passo que utiliza a história passada para prever eventos futuros. Um entendimento destas técnicas estatísticas especializadas é necessário para o processo ser bem feito. Os administradores que possuem *softwares* de gerenciamento de força de trabalho que automatizam o processo de previsão não podem pensar que isso apenas é suficiente; entender estes cálculos é tão crítico quanto possuir o *software*: não só para verificar a acurácia dos resultados como também, e talvez de forma mais importante, para explicar os números para a gerência. Ou seja, mesmo tendo as ferramentas, vale a pena entender os fundamentos do processo de previsão. (REYNOLDS, 2005)

Por isso, Hawkins *et al* (2001) dizem que os gerentes, supervisores e líderes de equipes dos *call centers* precisam receber – antes da central ser aberta e depois da sua completa implementação, de forma continuada – treinamentos específicos em diversas áreas, incluindo previsão da carga de trabalho e planejamento de horários da mão-de-obra.

Steckley, Henderson e Mehrotra (2006) afirmam que um dos objetivos dos gerentes das centrais de atendimento é identificar a quantidade mínima de agentes para atingir um bom nível de serviço. Mas essa tarefa não é fácil, em parte por causa dos erros no processo de previsão do volume de chamadas, que podem impactar na performance da atividade de planejamento.

Os autores seguem dizendo que muitas vezes é necessário aumentar a quantidade de atendentes para poder lidar com a hipótese da demanda ter sido subestimada devido a um erro de previsão. Se essa demanda por mais atendentes for emergencial, o reforço de contingente pode ter que ser conseguido através da terceirização de parte da mão-de-obra. Por tudo isso, é preciso identificar boas medidas para esses erros.

Para tal, Shen e Huang (2005b) dizem que é preciso conduzir um exercício de previsão com dados de fora da amostra (que serviu de base para a elaboração do modelo a ter seus erros mensurados). Duas medidas são usadas para estimar e comparar performances de previsão:

- o erro absoluto, que consiste no módulo da diferença entre os valores real e previsto para um determinado momento do tempo;

- e o erro relativo, normalmente apresentado na forma percentual, que consiste no erro absoluto dividido pelo valor real em um determinado momento do tempo.

Os autores enfatizam que essas duas medidas devem ser calculadas para cada dia e cada período de tempo; e definem algumas medidas capazes de sumarizar a performance global do processo de previsão, ao longo de vários períodos de tempo:

- o erro médio absoluto, que consiste na média dos erros absolutos de todos os momentos no tempo considerados no período de avaliação;
- o erro médio relativo, que consiste na média dos erros relativos de todos os momentos no tempo considerados no período de avaliação;
- a raiz do erro quadrático médio, que consiste na raiz quadrada da média dos erros absolutos (de todos os momentos no tempo considerados no período de avaliação) elevados ao quadrado.

A respeito da sua funcionalidade, uma previsão precisa – para ser completa – incluir não apenas o volume de ligações, mas também a grandeza tempo de atendimento. Para calcular a carga de trabalho e prever o tamanho da equipe e os planos de horários posteriormente, é necessária uma fotografia da carga de trabalho, que consiste no número de chamadas multiplicado pelo tempo médio de atendimento. (REYNOLDS, 2005)

O tempo médio de atendimento é um dos alvos atacados pelo método proposto por Shen e Brown (2005) para inferir acerca das curvas de regressões não-paramétricas, quando os erros seguem uma distribuição lognormal. O método é aplicado – entre outros propósitos – para modelar o padrão (dependente do tempo) do tempo médio de atendimento das chamadas de clientes recebidas no *call center* de um banco israelense.

Os autores mostram que os tempos de atendimento seguem uma distribuição lognormal e, então, modelam os mesmos por meio de uma função não-paramétrica dependente da hora do dia. O entendimento de tal comportamento variável destes tempos é essencial para compreender o ambiente operacional (dependente do tempo) de um sistema, e também para prever dinamicamente a sua carga de trabalho futura.

Reynolds (2005) sugere um processo de 4 passos para a realização de uma previsão de demanda (também válido para a grandeza tempo de atendimento) em um *call center*:

1. Analisar e ajustar a informação histórica – A chave para lidar com aberrações nos dados é primeiramente determinar as razões devido às quais elas ocorreram. Se for um incidente isolado, ou até que possa ocorrer novamente, mas cujo momento de ocorrência seja difícil de prever (como uma tempestade), é necessário normalizar os números para cima ou para baixo de forma a refletir volumes realistas; mas se for um evento que se repete, de forma previsível, os números precisam ser mantidos para que a previsão reflita o evento no futuro;

2. Prever as chamadas mensais – Para transformar os dados brutos em previsões para meses futuros, algumas abordagens estão disponíveis, como: estimação pontual, média simples, média ponderada, média móvel, amortecimento exponencial e séries temporais (incluindo análise de tendência e de sazonalidade);

3. Criar previsões diárias e para períodos de meia-hora – Para detalhar a previsão mensal em um escopo diário, é preciso calcular os fatores dos dias da semana – que podem apresentar sazonalidade – de forma a saber que percentual da carga de trabalho da semana cada dia representa. Depois é necessário repetir o processo para os padrões dos períodos do dia, para que sejam conhecidos os picos, vales e períodos médios em termos de volumes de ligações. É possível que seja preciso explicitar essa sazonalidade intradiária para cada um dos dias da semana.

4. Ajustar a previsão devido a outras influências do negócio – É preciso incorporar na previsão – com a devida antecedência – o efeito no volume de ligações que outras áreas da empresa são capazes de empreender como, por exemplo: o setor de *Marketing*, cujas decisões a respeito de vendas e promoções impactam tremendamente a demanda; o setor de Faturamento, que pode utilizar um novo formato de cobrança; o setor de Vendas, que pode conseguir novas contas de clientes; o setor de Logística, que pode mudar a embalagem e forma de envio dos produtos. Por tudo isso, é crítico que seja mantida uma comunicação regular com esses setores com potencial impacto sobre o volume e tempo de atendimento das chamadas.

A respeito dos picos de demanda, Hawkins *et al* (2001) afirmam que um número de situações pode ocorrer para causar a sua existência, e todas elas devem ser consideradas durante o desenvolvimento dos modelos de previsão. Os picos sazonais podem ocorrer em horizontes horários, diários, semanais, mensais e anuais e – para que o dimensionamento do sistema seja

feito de forma apropriada – os momentos em que e a magnitude com que eles ocorrem devem ser bem investigados ao longo de um período específico de tempo (pelo menos 1 ano).

Empresas que falham no planejamento dos picos e flutuações do volume de chamadas correm o risco de fornecer serviço inadequado e alienar clientes leais. Por isso, algumas empresas monitoram esse volume – para cada cliente – em tempo real e comparam com dados históricos. Isso permite que os gerentes saibam imediatamente se o volume de chamadas está excedendo a previsão e, se for o caso, reajam rapidamente aumentando a oferta de atendentes (ou, em caso contrário, mandando para casa agentes voluntários que queiram encerrar mais cedo o expediente). (ALEXANDER COMMUNICATIONS GROUP INC, 2005)

Shen e Huang (2005b) desenvolveram métodos para previsões dia-a-dia e intradiária (dinâmica) do volume de chamadas entrantes. A abordagem consiste em tratar os perfis intradiários de volume de chamadas como vetores multidimensionais de séries temporais.

A abordagem dos autores realiza inicialmente a redução da dimensionalidade através da Decomposição em Valores Singulares (SVD) da matriz de perfis intradiários históricos; e então aplica técnicas de séries temporais e de regressão. Tanto a dinâmica dia-a-dia quanto os padrões intradiários das chegadas de chamadas são levados em consideração pela abordagem.

Os métodos propostos pelos autores foram aplicados a um conjunto real de informações e mostraram melhorias substanciais em relação aos padrões existentes na indústria, em uma comparação com dados de fora da amostra utilizada para a elaboração dos modelos; além disso, são computacionalmente rápidos e, portanto, viáveis para uso em uma previsão dinâmica em tempo real.

Weinberg, Brown e Stroud (2006) propõem, em seu artigo, um modelo multiplicativo para prever taxas intradiárias de chegadas de chamadas ao *call center* de um banco comercial norte-americano. Métodos de amostragem de Monte Carlo para cadeias markovianas foram usados para estimar os estados latentes e os parâmetros do modelo; e um algoritmo seqüencial de Monte Carlo foi proposto para uma estimação e previsão dos parâmetros e taxas do modelo.

Adicionalmente, qualquer conhecimento anterior baseado em experiência passada do gerente pode ser incorporado ao modelo, assim como componentes observados somente *a posteriori*. Comparando as previsões (para o dia seguinte) fornecidas pelo modelo dos autores e por

modelos estatísticos clássicos, eles puderam perceber melhorias de até 25% por parte do seu modelo em relação aos padrões.

Shen e Huang (2005a) ilustram o uso da Decomposição em Valores Singulares (SVD) para visualização de informações, detecção de *outliers* (pontos que fogem do padrão) e redução de dados (implicando em uma representação parcimoniosa dos dados originais sem perda de muita informação) em um ambiente de *call center*. O trabalho descreve como os dados reduzidos podem ser usados para algumas análises estatísticas formais adicionais, incluindo o desenvolvimento de um modelo de previsão para o volume de chamadas.

Nele, os autores exploram um exemplo real, com dados coletados em um *call center* (do tipo *inbound*) de uma organização financeira norte-americana de grande porte. A SVD sugere que o volume diário de ligações e o padrão de chegadas diárias são dois componentes-chave, que são, então, usados para a construção de um modelo multiplicativo de previsão de curto prazo para o volume de ligações, contendo uma estrutura auto-regressiva de séries temporais que depende do dia da semana.

Os seguintes artigos também citam, descrevem, discutem e questionam a aplicação das técnicas de previsão de demanda em ambientes de *call centers*, merecendo destaque: BIANCHI, JARRET & HANUMARA (1998); ANDREWS & CUNNINGHAM (1995); BIANCHI, JARRET & HANUMARA (1993).

3.3. Teoria de filas

Dentro da ciência da administração, o termo Teoria de Filas consiste no corpo de conhecimento que lida com filas de espera. A teoria de filas foi concebida no início do século XX quando o engenheiro de telefonia dinamarquês A. K. Erlang começou a estudar o congestionamento e os tempos de espera que ocorriam no momento em que as ligações telefônicas eram completadas. Desde então, inúmeros modelos quantitativos têm sido desenvolvidos para ajudar pessoas de negócios a entenderem as filas, preverem seu comportamento em alguns casos particulares e fazerem melhores decisões acerca de como gerenciá-las. A teoria de filas é uma das mais antigas e mais bem pesquisadas áreas da ciência da administração. (RAGSDALE, 2001; BROWN *et al*, 2002)

Segundo Ragsdale (2001), muitos problemas de filas têm como foco determinar o nível de serviço que uma empresa deveria fornecer. Em muitos deles, a gerência tem algum controle sobre o nível de serviço fornecido. No entanto, isto pode ser caro, ou de fato um desperdício, se for mantido um número excessivo de servidores ociosos. Por outro lado, empregar um número pequeno de servidores mantém baixo o custo de fornecer o serviço, mas resulta em maiores tempos de espera para os clientes e sua maior insatisfação. Portanto, existe um *trade-off* entre o custo de fornecer o serviço e o custo de ter clientes insatisfeitos se o serviço for escasso.

Porém, de acordo com Hillier e Lieberman (1995), esta decisão normalmente é difícil de ser tomada, pelo fato de ser freqüentemente impossível prever acuradamente quando os clientes chegarão em busca do serviço e quanto tempo será necessário para o serviço ser prestado. Os autores também dizem que a teoria de filas, por si só, não resolve esse tipo de problema diretamente, ou seja, ela não busca atingir a meta de otimização na tomada de decisão; porém, ela contribui com informações vitais para estas decisões através da previsão de várias características da fila, como o tempo médio de espera, entre outras, e do comportamento do sistema. Na verdade, situações de fila que requerem tomada de decisão são encontradas em uma grande variedade de contextos. Por essa razão, não é possível apresentar um procedimento de tomada de decisão que seja aplicável a todas estas situações. Em outras palavras, todo problema individual tem suas próprias características especiais e, portanto, nenhum procedimento padrão pode ser indicado para se encaixar em todas as situações.

Hillier e Lieberman (1995) descrevem o processo básico assumido pela maior parte dos modelos de fila: clientes requerendo serviço são gerados ao longo do tempo por uma fonte de entrada. Estes clientes entram no sistema e se juntam a uma fila. Em certos momentos, um membro da fila é selecionado por alguma regra conhecida como disciplina da fila. O serviço requerido é então realizado para o cliente pelo mecanismo de atendimento, após o qual o cliente deixa o sistema de filas.

Para falar das características dos sistemas de filas, Ragsdale (2001) descreve a taxa de chegada. Em muitos sistemas de filas, os clientes chegam segundo um padrão randômico, ou seja, o número de chegadas que ocorrem em um determinado período de tempo consiste em uma variável aleatória. Normalmente, é apropriado modelar o processo de chegada em um sistema de filas como uma variável aleatória Poisson. Para usá-la, é necessário especificar um

valor para a taxa de chegada, representada por λ, que consiste no número médio de chegadas dentro de um intervalo de tempo. O tempo entre duas chegadas consecutivas é conhecido como tempo (ou intervalo) entre chegadas. Se o número de chegadas em um determinado período de tempo segue uma distribuição Poisson com média λ, pode ser demonstrado que o tempo entre chegadas segue uma distribuição exponencial com média $1/\lambda$. Na opinião do autor, a distribuição exponencial tem um papel chave nos modelos de fila, pois é uma das poucas distribuições "sem memória"; ou seja, o tempo até a próxima chegada não depende de quanto tempo se passou desde a última chegada. Segundo ele, a maior parte dos modelos de fila existentes assumem que as chegadas seguem uma distribuição exponencial.

Ragsdale (2001) apresenta também a taxa de serviço, dizendo que um cliente que chega a uma instalação de serviço passa uma quantidade de tempo (que pode ser nula) na fila (tempo na fila). A quantidade de tempo que o cliente gasta depois que o serviço inicia é conhecido como tempo de atendimento. De acordo com o autor, normalmente, é apropriado modelar este tempo como uma variável aleatória com formato exponencial. Para tal, uma taxa de serviço, representada por μ, precisa ser especificada. Ela consiste no número médio de clientes que podem ser atendidos por unidade de tempo. Nas distribuições exponenciais, pequenos tempos de atendimento têm mais probabilidade de ocorrência. Na prática, uma quantidade mínima de tempo é normalmente necessária para um serviço ser prestado. Como a distribuição exponencial abre possibilidade para essa variável assumir o valor zero (e também valores baixíssimos), ela tende a subestimar o tempo de atendimento real requerido para boa parte dos clientes. Esta distribuição também assume que tempos de atendimento altíssimos podem ocorrer (mesmo que de forma pouco freqüente), superestimando a variável. Na opinião do autor, os dois fenômenos acabam se compensando, permitindo que a distribuição exponencial forneça uma descrição razoavelmente acurada do comportamento do tempo de atendimento em muitos problemas do mundo real. Mas ele mesmo alerta que esta distribuição não consiste em um modelo adequado para o tempo de atendimento em todas as aplicações.

Ragsdale (2001) diz que, nas situações em que foi possível, pesquisadores desenvolveram equações para calcular várias características operacionais de modelos de filas específicos. Estas características operacionais podem ser úteis para os administradores que precisam tomar decisões acerca do *trade-off* entre o custo de fornecer o serviço e o custo de ter clientes insatisfeitos por causa do baixo nível de serviço.

Os principais cenários para os quais existem equações desenvolvidas na literatura estão relacionados a seguir:

- tempo entre chegadas e tempo de atendimento exponenciais, com 1 servidor;
- tempo entre chegadas e tempo de atendimento exponenciais, com mais de 1 servidor;
- tempo entre chegadas e tempo de atendimento exponenciais, com mais de 1 servidor e tamanho máximo de fila;
- tempo entre chegadas exponencial dependente do número de clientes no sistema (população finita), tempo de atendimento exponencial, com mais de 1 servidor;
- tempo entre chegadas exponencial, tempo de atendimento seguindo qualquer distribuição, com 1 servidor.

As equações desenvolvidas para estes cenários consideram a premissa da disciplina da fila ser FIFO (*first in, first out*) e podem ser encontradas em Ragsdale (2001).

Mas, segundo Hillier e Lieberman (1995) e Ragsdale (2001), para modelos de filas que fazem outras suposições (que não as de distribuição exponencial e Erlang), resultados analíticos úteis têm sido obtidos apenas em poucos casos. De fato, nem todos os modelos de fila têm equações fechadas para descrever as suas características operacionais.

3.3.1. Teoria de filas aplicada a *call centers*

Em um sistema de um *call center*, uma fila acontece quando não existe um agente disponível para atender um cliente, sendo este colocado em espera em uma fila virtual, da qual ele só sai quando um operador é alocado para atendê-lo ou quando ele abandona a chamada. Conforme observado por Brown *et al* (2002), no caso dos *call centers*, a fila virtual formada é invisível entre os clientes (que não vêem uns aos outros) e entre eles e os operadores que os servem.

No cenário de *call centers*, Araújo, Araújo e Adissi (2004) dizem que as disciplinas das filas, se bem gerenciadas, são fortes aliadas da área de planejamento e controle da produção dos *call centers*, que tem como meta atingir os resultados esperados com recursos, muitas vezes, escassos, tornando esta área cada vez mais importante nestas empresas. As disciplinas das

filas, quando bem administradas, podem trazer reduções significativas nos tempos de espera dos clientes que por elas passam.

O trabalho dos autores apresenta a bem sucedida experiência da implantação de uma disciplina de fila – diferente daquelas mais freqüentemente usadas – em uma empresa de telecomunicações, gerando economia e aumento da satisfação dos clientes por causa de melhorias no nível de serviço, no percentual de ligações atendidas e no tempo médio de espera. Para tanto, no lugar de filas simples não-priorizadas, foram feitas priorizações para o atendimento à demanda, considerando algumas variáveis associadas às ligações (seu tipo e, conseqüentemente, sua duração esperada). As filas "inteligentes", então, diferenciaram os diversos tipos de clientes, segmentando-os de acordo com o seu tipo de dúvida e a sua criticidade. Dessa forma, dentro de certos limites, clientes com dúvidas mais simples e com menor tempo esperado de atendimento, foram atendidos anteriormente – na frente de outros que chegaram antes ao sistema – ocasionando um melhor desempenho global do sistema.

Chassioti e Worthington (2004) descrevem um novo modelo para o gerenciamento das filas de um *call center* e apresentam os resultados da sua aplicação, que se revelam ser bem diferentes daqueles obtidos por modelos mais tradicionais. A abordagem analítica em questão é de fácil aplicação e oferece interessantes *insights* para o gerenciamento da atividade.

Algumas características das centrais de atendimento tornam difícil a aplicação de fórmulas analíticas da teoria de filas para a sua modelagem, incluindo: distribuições genéricas para o tempo de atendimento, taxas de chegada variando com o tempo, sobrecargas temporárias e abandonos. O novo modelo apresentado por Chassioti e Worthington (2004) consiste em uma abordagem prática capaz de incorporar a maioria dessas características.

Segundo Bapat e Pruitte Jr. (1998), as premissas adotadas nas análises baseadas nos modelos analíticos de teoria das filas são extremamente limitadas quando trabalhadas no contexto atual dos *call centers* porque:

- as chamadas recebidas são todas do mesmo tipo;
- a partir do momento em que uma chamada entra em uma fila, ela nunca a abandona – o que acaba, comumente, superestimando a necessidade de mão-de-obra e aumentado os custos de pessoal da empresa;

- os agentes atendem às chamadas fundamentados na disciplina onde o "**p**rimeiro que **e**ntra é o **p**rimeiro que **s**ai" (PEPS; ou FIFO, em inglês);

- cada operador atende todas as chamadas da mesma maneira.

Essas premissas raramente são válidas no ambiente em que as empresas de *call center* estão inseridas, já que, de acordo com os autores – dependendo da tolerância individual para esperar o seu atendimento – um cliente pode abandonar a ligação, caso seja colocado em uma fila. Além disso, os operadores normalmente diferem em relação às habilidades e também ao tempo de atendimento de uma chamada. Adicionalmente, as necessidades dos clientes são muito diferences e, algumas vezes, uma priorização – de forma que um serviço melhor possa ser prestado – se faz necessária.

Mas, Bapat e Pruitte Jr. (1998) acreditam que muitas empresas continuam apoiando as normalmente complexas tomadas de decisão acerca da alocação de recursos através dos modelos analíticos de teoria das filas, motivadas pelas relativas facilidade e rapidez da abordagem.

Muitas centrais de atendimento apresentam uma distribuição genérica (lognormal, por exemplo) para os seus tempos de atendimento, e não necessariamente uma distribuição exponencial negativa (BROWN *et al*, 2002). O formato exponencial é usado em grande parte da literatura da teoria de filas – tanto para tempo entre chegadas de clientes como para tempo de atendimento – pelo fato de existirem – para as situações onde esses tempos são considerados como variáveis que seguem uma distribuição exponencial – soluções analíticas para o estado estacionário do sistema.

Parte dos livros de teoria de filas (GROSS & HARRIS, 1985; KLEINROCK, 1975) trata de modelos estacionários e alguns (HILLIER & YU, 1981) apresentam tabelas e gráficos para o número de clientes no sistema e na fila.

No caso das centrais de atendimento existentes no mundo real, pelo menos a taxa de chegada de clientes varia com o tempo; essa variação é motivada por propagandas, horário de trabalho etc.. A fadiga dos atendentes pode gerar uma variação também no tempo de atendimento à medida que o tempo passa, mas esta é insignificante quando comparada à variação das taxas de chegada. As soluções encontradas na literatura para lidar com taxas de chegada variando

em função do tempo não são muito utilizáveis porque envolvem funções de Bessel, de difícil aplicação na prática. (CHASSIOTI & WORTHINGTON, 2004)

Entre os métodos de aproximação, os mais simples que podem ser usados são o *Simple Stationary Approximation* (SSA) (CHASSIOTI & WORTHINGTON, 2004) e o *Pointwise Stationary Approximation* (PSA) (GREEN & KOSELAR, 1991).

Na opinião de Chassioti e Worthington (2004), outros métodos de aproximação fazem-se, no entanto, necessários, porque o SSA e o PSA apresentam um desempenho pobre em casos importantes e também porque é necessário lidar com distribuições do tempo de atendimento diferentes da exponencial.

Os seguintes artigos também citam, descrevem, discutem e questionam a aplicação da teoria de filas (e outros métodos analíticos) a *call centers*, e merecendo destaque: WHITT (2006); BORST, MANDELBAUM & REIMAN (2004); FUKUNAGA *et al* (2002); KOOLE & MANDELBAUM (2002); GROSSMAN *et al* (2001); MANDELBAUM, SAKOV & ZELTYN (2001); DIJK (1997); e JENNINGS *et al* (1996).

3.4. Simulação

De acordo com Corrar e Theóphilo (2004), a história da simulação remonta aos jogos de guerra chineses, há 5.000 anos. Durante a Segunda Guerra Mundial, o matemático húngaro-americano John Von Neumann, em seu trabalho no Projeto Manhattan (da bomba atômica), criou um novo conceito, denominado Simulação de Monte Carlo.

Desde o final do século XX em diante, graças ao desenvolvimento dos recursos computacionais, esse método tem sido usado rotineiramente em diversas áreas, desde a simulação de fenômenos físicos complexos, como o transporte de radiação na atmosfera terrestre, até em causas mais populares, como na simulação do resultado de loterias. Pacotes conhecidos de *softwares* de simulação, como o Crystall Ball, @Risk, DecisionPro, Xcell, SLAM, Witness e MAP/1, servem ao campo dos negócios, ajudando os gestores na decisão sobre investimentos, melhoria de fluxos operacionais, políticas de estoque, manutenção etc.. (CORRAR & THEÓPHILO, 2004)

A simulação consiste em uma das técnicas mais gerais usadas em Pesquisa Operacional, que procura imitar a operação de um processo ou de um sistema da vida real ao longo do tempo,

com o auxílio de um modelo. Este modelo utiliza distribuições de probabilidade para gerar randomicamente vários eventos que ocorrem no sistema, tornando possível "comprimir o tempo" (dados relativos a meses e anos futuros podem ser obtidos em um pequeno período de tempo). Quer seja feita à mão ou através de um computador, a simulação envolve a geração de uma história artificial com o objetivo de obter inferências relativas às características operacionais do sistema real e melhorar o desempenho do mesmo através da comparação de alternativas e da análise de sensibilidade dos parâmetros do sistema, sem que ele precise ser efetivamente modificado ou até mesmo sem que necessariamente exista. (BANKS, CARSON & NELSON, 1999; HILLIER & LIEBERMAN, 1995; PIDD, 1998; SILVA et al, 1998; PEREIRA JR., 1999; CORRAR & THEÓPHILO, 2004; HERTZ, 1980)

Os modelos de simulação são analisados por métodos numéricos e não por métodos analíticos. Métodos analíticos aplicam raciocínios dedutivos da matemática para "resolver" o modelo. No caso dos modelos de simulação, que empregam métodos numéricos, os modelos são "rodados", ao invés de resolvidos. (BANKS, CARSON & NELSON, 1999; HILLIER & LIEBERMAN, 1995)

Costuma-se fazer uma analogia entre a simulação e a cirurgia clínica, já que ambas devem ser utilizadas como um último recurso, após esgotarem-se as demais alternativas de solução para um problema, muitas vezes mal-estruturado. Desta forma, esta ferramenta é tida como de valor incalculável para lidar com aqueles problemas onde técnicas analíticas são inadequadas, muito complexas ou inexistem; e é largamente utilizada em tais situações, onde grandes complexidades e incertezas (no caso da análise de risco) encontram-se presentes. (SALIBY, 1989; HILLIER & LIEBERMAN, 1995; SILVA et al, 1998; KLUNGLE & MALUCHNIK, 1997; RAGSDALE, 2001; CORRAR & THEÓPHILO, 2004; HERTZ, 1980)

Mas, segundo Saliby (1989) e Corrar e Theóphilo (2004), bons modelos de simulação podem ser caros, por exigirem longos e complicados processos de desenvolvimento (modelagem e programação) e de validação (tanto do modelo computadorizado quando do modelo conceitual). Adicionalmente, os resultados de um estudo de simulação estão sujeitos a variações (baixa precisão), além de não apontarem diretamente a solução ótima para o problema em estudo.

Toda simulação requer a construção de um modelo (com o qual serão feitos os experimentos), que deve começar de forma simples e ter sua complexidade aumentada aos poucos. Este

modelo é definido por um conjunto de relações lógico-matemáticas, descritas geralmente por um programa de computador. Dependendo do tipo de modelo, a simulação pode ser: determinística ou probabilística; estatística ou dinâmica; e discreta ou contínua. (SALIBY, 1989; PIDD, 1998)

Maiores informações sobre os três critérios de classificação da simulação podem ser encontradas em algumas fontes sobre o assunto. (SALIBY, 1989; SILVA et al, 1998; BANKS, CARSON & NELSON, 1999)

Já que muitas das suas restrições têm sido objeto de contínuos estudos por parte de especialistas (a partir dos quais muitos progressos vêm sendo registrados), a simulação tornou-se uma abordagem de estudo que vem sendo cada vez mais utilizada nas mais variadas áreas de conhecimento, inclusive por causa da sua flexibilidade: sistema de estoques e compras; planejamento financeiro (análise de risco); sistemas macroeconômicos; sistemas de transporte público; operações militares; e sistemas de atendimento (filas), tais como a operação de atendimento de um banco ou de uma central telefônica. (SALIBY, 1989)

3.5. Simulação em *call centers*

Chokshi (1999), Klungle e Maluchnik (1997), Hall e Anton (1998), Mehrotra e Fama (2003), Avramidis e L'Ecuyer (2005), Klungle (1999) e Bapat e Pruitte Jr. (1998) destacam alguns fatores recentes que contribuíram para o aumento da demanda pelo uso da ferramenta de simulação no setor de *call centers*:

- a crescente importância dos *call centers* para boa parte das corporações (devido ao rápido crescimento da era da informação e das comunicações e do aparato tecnológico), aumentando a necessidade de utilização de metodologias científicas de tomada de decisão e ferramentas para a sua gerência estratégica em substituição ao uso da intuição;

- a complexidade crescente do tráfico de chamadas unido a regras cada vez mais presentes de roteamento baseado em habilidades;

- a incerteza cada vez mais predominante nos problemas de decisão normalmente encontrados no gerenciamento operacional das centrais;

- rápidas mudanças nas operações e incremento nas atividades de reengenharia resultantes do aumento de fusões e aquisições, volatilidade do negócio, opções de terceirização e utilização de diferentes canais para atingir o consumidor (telefone, *e-mail*, *chat*);

- e a disponibilidade e o preço acessível dos computadores, aliados a uma gama de aplicações de simulação em *call center* disponíveis no mercado, cada vez mais intuitivas e fáceis de serem aprendidas e usadas.

Por isso, o assunto "simulação em *call centers*" é tão recente nos fóruns de discussão de simulação pelo mundo afora; e por isso também é que tem recebido bastante destaque nos últimos anos nos artigos, congressos e conferências de simulação. Por exemplo, a palestra de abertura da Winter Simulation Conference (um dos mais renomados encontros dos estudiosos de simulação) de 1999, ministrada pelo Dr. Jon Anton (Purdue University) teve como título "O uso da simulação na otimização de *call centers*".

A simulação, de acordo com Mehrotra (1997), modela explicitamente a interação entre chamadas, rotas e agentes, assim como a aleatoriedade das chegadas de chamadas individuais e dos tempos de atendimento. Através do uso da simulação, gerentes e analistas traduzem dados brutos de *call centers* (previsão de chamadas, distribuições de tempos de atendimento, horários e habilidades de agentes, vetores de roteamento de chamadas etc.) em manuseável informação sobre níveis de serviço, abandono de clientes, utilização de agentes, custos e outras importantes medidas de performance em um *call center*.

Segundo Chokshi (1999) e Klungle e Maluchnik (1997), a utilização da simulação para auxiliar a tomada de decisões no gerenciamento de *call centers* permite os seguintes benefícios:

- visualizar processos futuros e ser usada como uma ferramenta de comunicação;

- validar premissas de processos antes de sua implementação;

- analisar detalhadamente o impacto de mudanças (análise de cenários);

- prever necessidades agregadas de recursos e programar a força de trabalho;

- mensurar indicadores de performance;

- estimar impactos nos custos e economias.

A primeira utilidade da simulação em um *call center*, segundo Hall e Anton (1998), é a de avaliação, quando se pode verificar "onde o *call center* está". A pergunta-chave é "quão eficiente e efetivo é a operação atualmente?". O objetivo da avaliação é estabelecer uma linha de partida (e de referência) para a mudança.

A avaliação situa o *call center* específico dentro das métricas da indústria, analisa o *status quo* e fornece uma fotografia para o futuro, que pode envolver um completo realinhamento do *call center* e de sua missão para apoiar a direção estratégica da empresa. A simulação computacional é uma ferramenta para se estabelecer essa linha de partida e de referência. (HALL & ANTON, 1998)

O artigo dos últimos autores mostra um estudo de caso ilustrativo de avaliação de uma grande central de atendimento de serviço de utilidade pública nos EUA. Klungle (1999) apresenta um caso onde uma aplicação em um *call center* de reclamações de seguros é utilizada para mostrar a efetividade da simulação ao avaliar desenhos de centrais de atendimento.

O estudo de simulação realizado por Gulati e Malcolm (2001) em um *call center* (do tipo *outbound*) de um grande banco mostrou que a sua gerência estava correta, quando presumira que o *software* usado para fazer a programação das chamadas – baseado em uma heurística – não estava fornecendo um nível de serviço adequado.

Mehrotra, Profozich e Bapat (1997) falam da segunda utilidade da simulação em um *call center*: sem atrapalhar a operação da central e sem impactar no seu orçamento, a simulação permite o entendimento rápido e acurado de como será a performance operacional do *call center* diante de certos cenários (baseados em modificações causadas por iniciativa externa ou gerencial como, por exemplo, uma nova tecnologia, uma nova estratégia de negócio ou um aumento da carga de trabalho), antes que qualquer mudança seja efetivamente feita.

Segundo os autores, dessa forma, algumas questões podem ser respondidas, entre outras:

- Qual o impacto de um *overflow* de ligações?

- Quais são os *trade-offs* envolvidos no ato de priorizar as chamadas de clientes especiais?

- O serviço pode ser melhorado se forem colocados despachantes para fornecer informações básicas aos clientes?

No tocante à análise de cenários, o trabalho de Gulati e Malcolm (2001) usou a simulação para comparar a performance de três diferentes abordagens de programação de chamadas (heurística, otimização por lotes diários e otimização dinâmica – de hora em hora), revelando existir oportunidades de melhoria no processo do *call center* (do tipo *outbound*) do banco estudado. Modelos de simulação computacional revelaram-se ser a única maneira prática para desenvolver as estimativas comparativas, por causa da escassez de disponibilidade de dados comparativos entre as duas primeiras abordagens e pelo fato da terceira abordagem nunca ter sido implantada. Através destes modelos, foi possível imitar o processo através do qual as chamadas são realizadas a partir de uma lista produzida por cada um dos três algoritmos de programação. Os *outputs* do modelo forneceram uma maneira de verificar a performance do sistema em relação às metas da administração e mostraram que as abordagens não heurísticas produziriam resultados superiores, mas não durante todo o período do dia.

Em relação ao mesmo tema, Miller e Bapat (1999) descrevem como a simulação foi usada para projetar o ROI da compra e utilização de uma nova tecnologia de roteamento de chamadas para 25 centrais de atendimento. Com investimentos de US$ 17 milhões e custos anuais de operação de US$ 8 milhões para a nova tecnologia, era preciso determinar se ela provocaria benefícios suficientes (através de redução nos custos dos troncos, aumento da produtividade dos operadores e possibilidade de atender mais chamadas) para garantir a sua implementação em nível nacional. Foram conduzidas simulações diárias com pequenos e grandes volumes (o que acarretou grande variação nos resultados) e com outras análises de cenários para facilitar as decisões de planejamento. Um protótipo da nova tecnologia foi usado na tentativa de comparação de resultados, mas foi a simulação que tratou a questão da melhor maneira. (MILLER & BAPAT, 1999)

Ainda dentro do assunto referente à segunda utilidade da simulação (análise de cenários), Lam e Lau (2004) descrevem um esforço de reestruturação de uma empresa de Hong Kong, que fornece serviços de assistência técnica para equipamentos de escritório e computadores. Estando disponíveis muitas oportunidades de melhorias no processo, a abordagem de simulação foi usada para explorar as diferentes opções e para avaliar os resultados da reestruturação dos *call centers* existentes. A análise dos resultados simulados confirmou que a

grande oportunidade de melhoria consistia na junção dos recursos existentes em uma única central. De forma geral, segundo os autores, a gerência de um *call center* pode – assegurada pelos resultados da simulação – avaliar diferentes benefícios – tangíveis e intangíveis – antes de implementar o plano de reestruturação.

Outro exemplo na mesma linha pode ser encontrado no estudo apresentado por Saltzman e Mehrotra (2001), onde foi aplicada a ferramenta de simulação em uma grande empresa de *software* que pretendia visualizar o funcionamento operacional do seu *call center* antes do lançamento de um novo programa de serviço de suporte pago. A idéia consistia em verificar se seria cumprida a meta de os clientes pagantes terem que esperar menos de 1 minuto para serem atendidos. Os gerentes também queriam saber qual seria o impacto do novo programa sobre o serviço oferecido à base regular de clientes não pagantes. Com o uso da simulação, foi possível prever o comportamento do sistema e, assim, tomar as medidas necessárias para garantir o sucesso do programa.

Segundo Mehrotra, Profozich e Bapat (1997), Gulati e Malcolm (2001), Bapat e Pruitte Jr. (1998) e PARAGON (2005), um modelo de simulação pode (e tem sido cada vez mais usado para tal) – além de normalmente permitir gráficos e animações – contemplar alguns aspectos críticos dos *call centers* receptivos modernos de todos os tamanhos e tipos, entre outros:

- nível de serviço específico;
- flexibilidade na distribuição dos tempos entre chegadas e de atendimento;
- consolidação de centrais;
- roteamento baseado em habilidades;
- tipos múltiplos de chamadas;
- filas simultâneas;
- padrões de abandono de chamadas;
- retorno de ligações;
- *overflow* e transbordo;

- priorização de filas;
- transferência de chamadas e teleconferências;
- preferências, proficiência, tempo de aprendizado e esquema de horário dos operadores.

Os autores acrescentam que os *outputs* do modelo podem vir na forma de tempo de espera, taxa de abandono de chamadas (ambos com possibilidade de diferenciação por tipos de chamada) e nível de utilização dos operadores (com possibilidade de diferenciação por tipos de operador). E devido à adequação desta abordagem às características reais e complexas dos *call centers*, a simulação pode tornar mais confiável o seu dimensionamento e gerenciamento.

De acordo com Mehrotra, Profozich e Bapat (1997), Steckley, Henderson e Mehrotra (2005), PARAGON (2005), Mehrotra (1997), Klungle e Maluchnik (1997), Pidd (1998) e Tanir e Booth (1999), os métodos tradicionais mais usados no gerenciamento e dimensionamento de um *call center* (estimativas intuitivas, cálculos improvisados, planilhas e modelos teóricos de filas Erlang) estão se tornando significativamente limitados por causa da variabilidade das chegadas de chamadas, das rotas e do tempo gasto nas ligações, das habilidades e prioridades dos operadores, da heterogeneidade das chamadas e da interação entre elas e os troncos, da dinâmica de abandono das chamadas, das tendências recentes (como o roteamento baseado em habilidades, canais eletrônicos e manuseio interativo de chamadas) e, de uma forma geral, da sofisticação e complexidade cada vez mais presentes nos sistemas de *call centers*. Por exemplo, os modelos analíticos comumente assumem que a chegada dos clientes consiste em um processo Poisson quando, na verdade, os dados de *call centers* rejeitam esta premissa freqüentemente. Adicionalmente, planilhas e modelos Erlang superestimam a quantidade de atendentes, além de terem pouca precisão para centrais com atendimento diferenciado.

A simulação estende a capacidade das ferramentas analíticas e operacionais e consiste em uma ferramenta superior quando não está disponível nenhum modelo teórico tratável que forneça uma representação razoável do sistema sob estudo e as médias não são suficientes, a acurácia é importante, a operação é detalhada, a demanda varia muito, gargalos e necessidades de mudanças no desenho dos processos precisam ser identificados e/ou uma animação é necessária para a melhor comunicação de uma modificação para a diretoria. As recentes tendências da indústria demandam abordagens mais sofisticadas e a simulação fornece as técnicas necessárias para se adquirir *insights* sobre essas novas tendências e ajuda a

dar forma aos seus desenhos atuais e futuros, consistindo no único método de análise apto a modelar um *call center* eficiente e acuradamente, através de uma abordagem muito mais prática, flexível em termos de *inputs* e *outputs*, e capaz de permitir a inclusão de detalhes importantes, refletir muito melhor a realidade (sem grandes necessidades de simplificações, como nos modelos teóricos), viabilizar um melhor e mais profundo entendimento acerca dos processos da central e gerar resultados muito mais robustos a respeito da performance do *call center*, permitindo sua otimização de forma bem mais confiável. (PARAGON, 2005; RILEY, 2005; MEHROTRA, 1997; KLUNGLE & MALUCHNIK, 1997; TANIR & BOOTH, 1999; SALIBY, 1989; HILLIER & LIEBERMAN, 1995; HERTZ, 1980; MEHROTRA, PROFOZICH & BAPAT, 1997; BAPAT & PRUITTE JR., 1998; CHOKSHI, 1999; KLUNGLE, 1999; WORTHINGTON & WALL, 1999; RAGSDALE, 2001; MEHROTRA & FAMA, 2003)

Naturalmente, o emprego da simulação não pressupõe a completa eliminação da participação de outras abordagens no processo de dimensionamento de uma central, podendo inclusive ocorrer uma integração entre diversas técnicas, em alguns casos, como no caso retratado por Barboza *et al* (2003).

Os autores propõem uma solução para a elaboração e designação de horários de atendentes em uma central telefônica de atendimento 24 horas terceirizada, envolvendo três fases: determinação (através de uma simulação da central) do número necessário de atendentes para cada horário do dia, sendo fixado um certo nível de serviço (que influencia diretamente o faturamento da empresa terceirizada); verificação (através de Programação Inteira – algoritmo *Branch and Bounds*) dos melhores horários dos atendentes de forma a minimizar os custos da empresa (salários), em função do número de atendentes em cada categoria (3, 4 e 6 horas de trabalho diárias) e dos resultados da primeira fase; e designação (através do algoritmo de *Matching* de Peso Máximo) do horário de cada atendente, no sentido de maximizar a sua satisfação (medida de acordo com questionários de preferência de horários) e atender os resultados da segunda fase.

Uma semana real específica é comparada com esta mesma semana simulada, contemplando o número de chamadas ocorrido a cada meia hora, o número de atendentes sugeridos pelo modelo para atingir um nível de serviço pré-determinado, o número de atendentes efetivamente utilizado pela empresa naquele período durante aquela semana e o nível de

serviço efetivamente atingido pela empresa a cada meia hora. Essa comparação revela que a empresa poderia escalonar de maneira melhor os seus funcionários, já que em alguns períodos, houve excesso de mão-de-obra e em outros houve falta, acarretando níveis de serviço maiores e menores, respectivamente, do que o estipulado (ou desejado). Dessa forma, a empresa poderia usar um menor número de funcionários (e reduzir seus custos) e atingir um nível de serviço melhor (e estável), o que aumentaria diretamente seu faturamento. Outra vantagem seria a maior satisfação em relação à preferência de horários que seria alcançada entre os atendentes. (BARBOZA *et al*, 2003)

Mehrotra e Fama (2003) e Klungle (1999) enxergam tendências futuras capazes de impactar a simulação de *call centers*:

- a complexidade operacional, que vai continuar a crescer – na forma de mais filas, maior variedade na escala dos operadores e diversidade de combinação de habilidades e regras de roteamento – obrigando os analistas a criarem modelos mais ricos;

- o surgimento de cada vez mais *softwares* de simulação especializados em *call center*, cuja importância tende a acompanhar o papel que a simulação assumirá no processo de redesenho das centrais, necessário para lidar com a sua nova complexidade;

- o entendimento maior – por parte dos executivos – que os *call centers* são componentes centrais na cadeia de valor do cliente, disparando um desejo de entender os riscos inerentes de qualquer configuração operacional e a conseqüente melhoria na qualidade dos dados coletados e na acurácia dos parâmetros (como distribuição do tempo entre chegadas, tempo de atendimento, tempo de espera, taxa de abandono e outros), encerrando resultados mais robustos.

Os seguintes artigos também citam, descrevem, discutem e questionam a aplicação da simulação a *call centers*, merecendo destaque: ELANGO (2006); MARKT (2005); ATLASON & EPELMAN (2002); GROSSMAN *et al* (2001).

3.5.1. *Softwares* de simulação para *call centers*

Segundo Mehrotra (1997), ao longo dos últimos anos, pacotes de simulação disponíveis comercialmente tem sido desenvolvidos especificamente para *call centers*, permitindo

gerentes e analistas – sem extensas habilidades matemáticas nem de programação – construir e rodar modelos de *call center* rapidamente.

3.5.1.1. Arena Contact Center

Devido ao sucesso do uso da simulação em ambientes de *call center*, a Systems Modeling Corporation em parceria com a Onward (consultoria especializada em análise e melhoria da operação de *call centers*), desenhou um sistema de simulação – chamado **Arena Contact Center** (antigo Call$im) – construído a partir do *software* Arena, especialmente para o desenvolvimento de modelos para *call centers*, com animação e apresentação de estatísticas em tempo real e também na forma de relatórios. (PARAGON, 2005)

De acordo com a fonte, o *software* possibilita criar um *call center* virtual no computador, para avaliação de investimento, nível de serviço, turnos, etc. sem riscos. Ele apresenta grandes vantagens sobre outros produtos de simulação, já que é simples, amigável, focado e usa a terminologia de *call centers* sempre que possível, com o objetivo de tornar intuitivos a construção e o uso do modelo. A maioria das ferramentas de simulação é de propósito geral, o que requer muito mais profundidade de conhecimento para criar um modelo de trabalho. Pelo fato de poucas pessoas atualmente disporem de semanas ou meses para aprender um novo *software*, usar essas ferramentas genéricas e sofisticadas não é muito prático para um gerente de *call center*. O Arena Contact Center permite a geração do modelo inicial em poucas horas.

Klungle (1999) acrescenta que, além de ser mais flexível em termos de modelagem, o *software* fornece uma riqueza de informações sobre performance do sistema, incluindo nível de serviço, taxas de abandono e de utilização dos operadores e capacidade do sistema, tornando-se extremamente útil para gerentes, planejadores e analistas operacionais e financeiros de *call centers* e para consultores na área.

Algumas perguntas que o *software* pode responder, segundo (PARAGON, 2005), são:

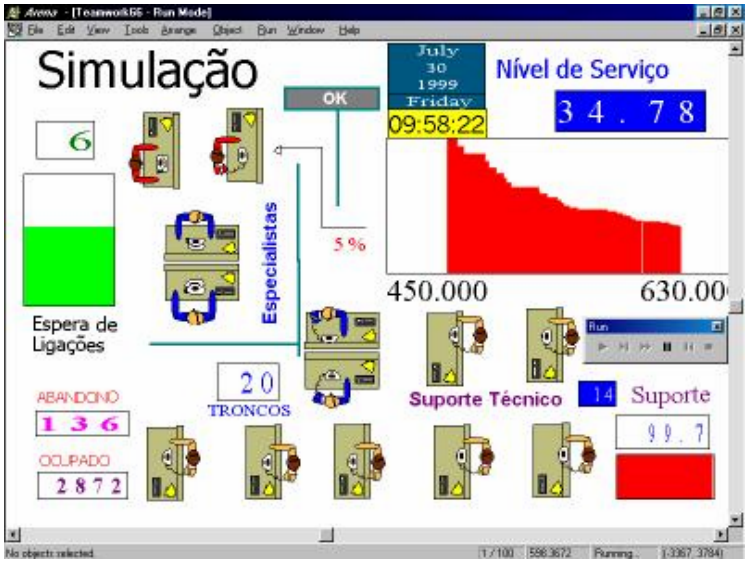
- Quantos atendentes são necessários para atingir o nível de serviço desejado?
- Como é possível atender a picos de demanda sem terceirização?
- Como os novos roteamentos de chamadas vão impactar a operação?

- Qual o impacto de transferir ligações, em diversos centros? E com a adoção de uma nova URA?

- Qual o impacto de adotar o conceito de "cliente preferencial"?

De acordo com a mesma fonte, a modelagem é 100% gráfica, em ambiente semelhante a um fluxograma, o que pode ser conferido na figura 4 a seguir. Ela é facilitada pelo módulo Input Analyzer, que analisa os dados de entrada fornecidos pela central telefônica e ajusta a melhor equação matemática para a modelagem.

Figura 4 – Modelagem de um *call center* usando o Arena Contact Center



Fonte: www.paragon.com.br

Na figura 4 anterior, os agentes (especialistas e de suporte técnico) podem ser visualizados ao mesmo tempo em que alguns indicadores dinâmicos (ligações em espera, abandonadas, nível de serviço etc.) vão sendo atualizados.

Ainda segundo a fonte, os seis passos básicos para a construção de um modelo usando a ferramenta, de forma fácil e rápida, são:

- Definição dos grupos de troncos e dos parâmetros de configuração do *call center*;

- Especificação do esquema de horário dos operadores;

- Definição dos grupos de operadores;

- Descrição do roteamento das chamadas;
- Entrada dos padrões de chegada das chamadas;
- Descrição dos tipos e quantidades de chamadas e do comportamento de abandono.

Wallace e Saltzman (2005) descrevem a modelagem de um *call center* com roteamento baseado em habilidades usando dois métodos distintos de programação em simulação: a linguagem C e o pacote Arena. Depois de salientarem os componentes proeminentes de cada método no processo de modelagem do *call center*, os autores concluem o artigo com uma comparação entre os prós e contras do uso de cada abordagem de programação dentro do contexto em questão.

Estas foram algumas atividades que o Arena Contact Center foi capaz de realizar, ao lidar com problemas reais (dentro de cenários diferentes), de acordo com (PARAGON, 2005):

- mensurar o impacto nos níveis de serviço de um aumento de 10% no volume de chamadas e o *payback* de um treinamento multifuncional para parte dos operadores;
- verificar como uma redução de 20% no orçamento do *call center* poderia afetar os níveis de serviço (no caso de redução de pessoal) e como o horário de trabalho poderia ser diminuído para essa redução no orçamento ser atingida;
- discutir o quanto custaria para o *call center* a terceirização de parte das chamadas e qual seria o ponto de equilíbrio em termos de aumento do número de operadores.

Segundo a mesma fonte, o *software* ainda pode ser integrado ao CallPeople, que consiste em uma ferramenta de otimização e programação de escala de trabalhos de atendentes, desenvolvida especificamente para *call centers* pela PARAGON/SEED, levando em consideração todas as características da CLT, sendo um produto totalmente adaptado ao mercado brasileiro. Tal integração objetiviza realizar o dimensionamento e o seqüenciamento de escalas para centrais de atendimento.

Os principais clientes do *software* pelo mundo, de acordo com a mesma fonte, são: Accenture; AT&T; Bank of América; BankBoston; Barclays Bank; Bell Canadá; Bell South; Delta Airlines; Ernst & Young; Federal Express; First USA Bank; ING Bank; Intuit; Lucent

Technologies; Microsoft; Oracle Corporation; Purdue University; Siemens; Southwestern Bell; UPS - United Parcel.

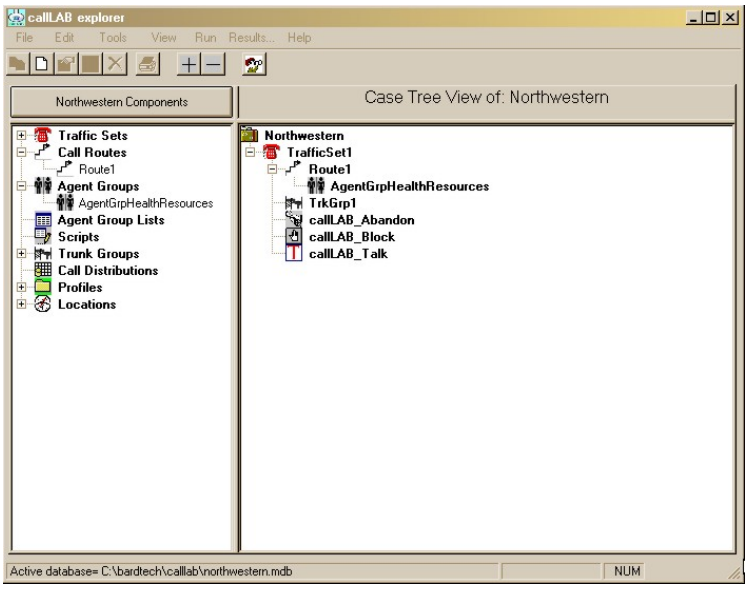
3.5.1.2. callLAB

O **callLAB** é uma ferramenta de planejamento, que emprega a tecnologia de simulação chamada-a-chamada, desenvolvida pela BARD Technologies expressamente para abordar e determinar o impacto de diversas características dos *call centers* modernos – roteamento de rede, múltiplas centrais, agentes com múltiplas habilidades, equipamentos sofisticados para melhorar a performance da central – sendo capaz de simular as complexidades de virtualmente qualquer operação de *call center*. (BARD TECHNOLOGIES INC, 2005)

A mesma fonte diz que o *software* pode considerar *inputs* – a serem introduzidas no modelo em uma tela como a ilustrada na figura 5 a seguir – como, entre outros:

- volume de chamadas
- número de troncos
- características da equipe
- tempo de atendimento
- regras de roteamento de chamadas;
- comportamento dos clientes quanto às chamadas.

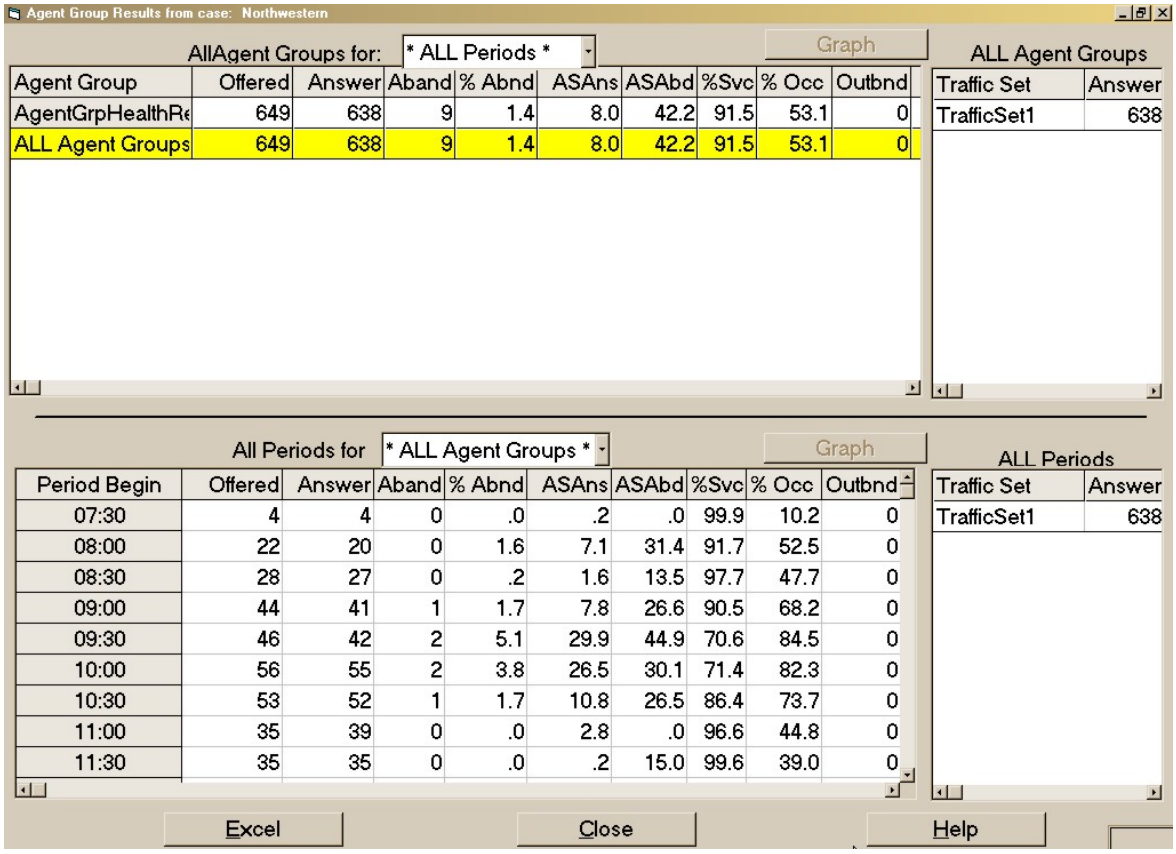
Figura 5 – Tela de entrada de variáveis do callLab



Fonte:www.bardtech.com

Ele é capaz de obter acuradamente resultados acerca do nível de serviço, ocupação dos agentes, utilização dos troncos etc. (conforme pode ser observado na figura 6 a seguir). O callLAB também pode ajudar a determinar o impacto de idéias como roteamento baseado em habilidades, roteamento de rede e mistura de chamadas – antes da sua implementação (e, conseqüentemente, antes dos equipamentos serem adquiridos, dos atendentes serem contratados ou demitidos e dos clientes ficarem insatisfeitos). (BARD TECHNOLOGIES INC, 2005)

Figura 6 – Tela de resultados do callLab



Agent Group	Offered	Answer	Aband	% Abnd	ASAns	ASAbd	%Svc	% Occ	Outbnd
AgentGrpHealthRe	649	638	9	1.4	8.0	42.2	91.5	53.1	0
ALL Agent Groups	649	638	9	1.4	8.0	42.2	91.5	53.1	0

Traffic Set	Answer
TrafficSet1	638

Period Begin	Offered	Answer	Aband	% Abnd	ASAns	ASAbd	%Svc	% Occ	Outbnd
07:30	4	4	0	.0	.2	.0	99.9	10.2	0
08:00	22	20	0	1.6	7.1	31.4	91.7	52.5	0
08:30	28	27	0	.2	1.6	13.5	97.7	47.7	0
09:00	44	41	1	1.7	7.8	26.6	90.5	68.2	0
09:30	46	42	2	5.1	29.9	44.9	70.6	84.5	0
10:00	56	55	2	3.8	26.5	30.1	71.4	82.3	0
10:30	53	52	1	1.7	10.8	26.5	86.4	73.7	0
11:00	35	39	0	.0	2.8	.0	96.6	44.8	0
11:30	35	35	0	.0	.2	15.0	99.6	39.0	0

Traffic Set	Answer
TrafficSet1	638

Fonte:www.bardtech.com

3.5.1.3. SimACD

Segundo (PORTAGE COMMUNICATIONS INC, 2005), com o **SimACD**, produto da Portage Communications, é possível testar diferentes números de agentes e filas de chegada antes da sua real implantação. Após a entrada – em uma tela como a ilustrada na figura 7 a seguir – dos volumes esperados de chamadas, das características do tempo de atendimento, do número de agentes e filas planejados para cada período do dia de trabalho, do comportamento de abandono e de "re-tentativa" por parte dos clientes e de características de *overflow* da central, o *software* permite:

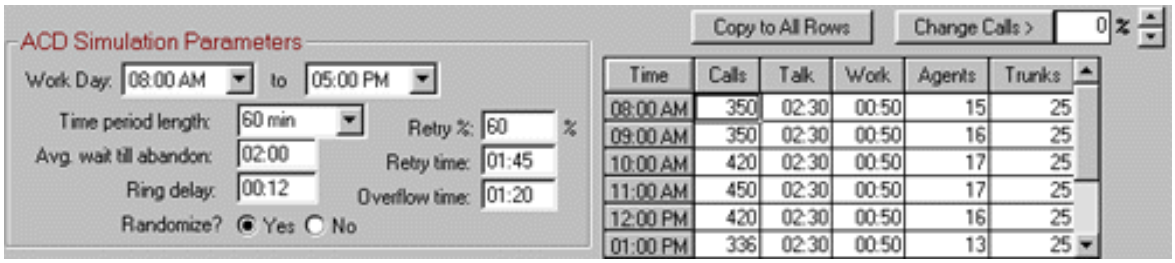
- que se assista a execução da simulação minuto-a-minuto, chamada-a-chamada, para cada dia durante a semana;

- que se observe o resultado das chegadas de chamadas, sendo atendidas, ficando na fila, recebendo o sinal de ocupado, sendo abandonadas, "re-tentadas" e causando *overflow*;

- computar e fazer gráficos acerca das estatísticas de performance;

- re-rodar simulações com diferentes números de agentes e filas até achar um cenário satisfatório em termos de níveis de serviço, carga de trabalho dos agentes e aspectos financeiros do *call center*.

Figura 7 – Tela de parâmetros do SimACD



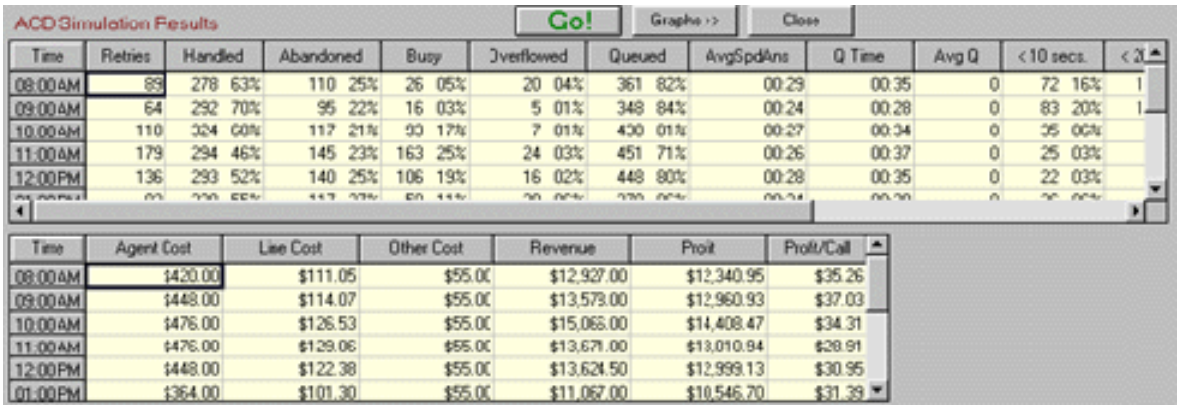
Fonte: www.portagecommunications.com

De acordo com a fonte, o SimACD comprime um dia de tráfego de chamadas em um curto período de tempo, representando cada segundo do dia como poucos micro-segundos de tempo no computador. Chamadas chegam randomicamente a um *call center* virtual criado no computador e são atendidas por agentes, enfileiradas, encontram sinal de ocupado, transbordam ou são abandonadas dependendo dos parâmetros, volumes de chamadas e combinações de agentes e troncos definidos para a central. Ele funciona como um *call center* experimental, que reage à chegada randômica de chamadas modeladas a partir da central real.

Ao final da simulação, é possível verificar quantas chamadas atingiram cada *status* (atendida, abandonada etc.), além do comprimento das filas, velocidade média para atender uma chamada e níveis de serviço reais para cada atendente, como pode ser observado na figura 8 a seguir. (PORTAGE COMMUNICATIONS INC, 2005)

Segundo a mesma fonte, alguns dos principais *call centers* clientes do SimACD são: Help Desk Institute, IBM Global Services, AT&T Call Center Solutions, BASF Corporation, British Airways, MCI Call Center Management, Ticket Máster, Pizza Hut, entre outros.

Figura 8 – Tela de resultados do SimACD



ACD Simulation Results

Go! | Graphs -> | Close

Time	Retries	Handled	Abandoned	Busy	Overflowed	Queued	AvgSpdAns	Q Time	Avg Q	<10 secs.	< 20
08:00AM	89	278 63%	110 25%	26 05%	20 04%	361 82%	00:29	00:35	0	72 16%	1
09:00AM	64	292 70%	95 22%	16 03%	5 01%	348 84%	00:24	00:28	0	83 20%	1
10:00AM	110	324 60%	117 21%	93 17%	7 01%	430 01%	00:27	00:34	0	35 06%	
11:00AM	179	294 46%	145 23%	163 25%	24 03%	451 71%	00:26	00:37	0	25 03%	
12:00PM	136	293 52%	140 25%	106 19%	16 02%	448 80%	00:28	00:35	0	22 03%	

Time	Agent Cost	Line Cost	Other Cost	Revenue	Profit	Profit/Call
08:00AM	$420.00	$111.05	$55.00	$12,927.00	$12,340.95	$35.26
09:00AM	$448.00	$114.07	$55.00	$13,573.00	$12,960.93	$37.03
10:00AM	$476.00	$126.53	$55.00	$15,066.00	$14,408.47	$34.31
11:00AM	$476.00	$129.06	$55.00	$13,671.00	$13,010.94	$28.91
12:00PM	$448.00	$122.38	$55.00	$13,624.50	$12,999.13	$30.95
01:00PM	$364.00	$101.30	$55.00	$11,067.00	$10,546.70	$31.39

Fonte: www.portagecommunications.com

3.5.1.4. ServiceModel

De acordo com (BELGE SIMULAÇÃO, 2005), o **ServiceModel** é um simulador confeccionado pela PROMODEL e destinado ao setor de serviços em geral, para solucionar questões como:

- Planejamento de capacidade;
- Programação de serviços e equipes;
- Fluxos complexos de atendimento a clientes;
- Racionalização máxima mantendo níveis de atendimento.

Segundo a fonte, o *software* é usado em programas tanto de graduação quanto de pós-graduação, notoriamente em cursos de administração de empresas, os quais buscam enfatizar a melhoria no gerenciamento de serviços, inclusive de *call centers*, que estão entre as suas principais aplicações. Dentro deste setor, os principais clientes estão na área financeira, destacando-se American Express (DECKER, 1999) e Chase Manhattan Bank.

Na 1ª Conferência Internacional de Simulação e Usuários Promodel, realizada em São Paulo, em 1999 (Innovation-99), realizada pela Belge Simulação, William Carrier, da PROMODEL (EUA), ministrou uma palestra cujo título era "ServiceModel melhorando as centrais de atendimentos - '*call centers*'".

3.5.1.5. Simul8

O **Simul8** é o *software* de simulação que possui o maior número de usuários no mundo. Foi desenvolvido pela Simul8 Corporation, com o objetivo de atingir um custo não impeditivo (para pequenas empresas), mas contendo todas as funcionalidades dos melhores *softwares* do mercado. O *software* em questão não é específico para *call centers*, mas o setor consiste em uma das principais sugestões de aplicação dadas pelo fabricante; além disso, o programa foi usado como motor para a aplicação ccProphet (que será descrita na seção 3.5.1.6) e apresenta vários projetos desenvolvidos na área de *call centers*. (SIMULATE, 2005)

O setor de *call centers*, segundo a mesma fonte, é um segmento onde a simulação pode ser plenamente aplicada. A simulação em um *call center* pode determinar qual é a melhor configuração de uma ilha de atendimento, qual é o número ideal de PAs (posições de atendimento) necessárias em uma certa hora do dia para garantir um determinado nível de serviço, quais equipamentos de automação de *call centers* devem ser utilizados etc..

Alguns tipos de centrais já modeladas em projetos desenvolvidos com o *software* foram: *call center* receptivo com 2 ilhas, *call center* ativo e receptivo de chamadas transbordadas, *call center* receptivo com roteamento baseado em habilidades, entre outros. As principais empresas cujos *call centers* utilizam o Simul8 são: One Call - Call Center, Sky – TV, Sher Marketing e Softway - Contact Center. (SIMULATE, 2005)

3.5.1.6. ccProphet

De acordo com (NOVASIM, 2005), o **ccProphet** é uma aplicação de *software* baseada em simulação projetada especificamente para análise de *call centers*. Com ela, a altamente sofisticada tecnologia de simulação computacional fica acessível à indústria de *call centers*. Usando simulação de eventos discretos com uma interface amigável, remove o passo na curva de aprendizagem tipicamente presente no uso de *softwares* de simulação e também a necessidade de contratar caros especialistas para obter todos os benefícios desta tecnologia. O ccProphet foi desenvolvido pela NovaSim e usa o Simul8 como motor de simulação. Com ele, é possível:

- gerenciar filas que chegam à central;
- balancear múltiplos modos de entrada de chamadas;

- planejar equipes com múltiplas habilidades;
- analisar indicadores referentes aos clientes e estabelecer níveis de serviço;
- avaliar novas tecnologias.

3.5.1.7. ContactCenters

Buist e L'Ecuyer (2005) apresentam a arquitetura geral e os componentes principais (e suas interações) de um modelo desenvolvido por eles em Java para a confecção de simuladores de *contact center*. **ContactCenters** suporta centrais com múltiplas habilidades e processos de chegada, roteamento e políticas de discagem arbitrários e complexos.

O programador pode alterar a lógica da simulação de várias maneiras, sem modificar o código básico do modelo. ContactCenters também pode operar em conjunto com outros modelos para fins de, por exemplo, otimização e análise estatística. Performance, flexibilidade e extensibilidade foram as principais metas do seu projeto e implementação.

3.5.1.8. Excel Plataform for Simulating Call Centers (EPSCC)

A performance operacional de *call centers* é freqüentemente medida através de tempos na fila e taxas de abandono de clientes e, por isso, os gerentes precisam realmente entender como as políticas de gerenciamento e os fatores estocásticos afetam estas estatísticas de performance. A simulação é um excelente veículo para examinar essas relações, mas uma carência na habilidade de programação pode ser uma barreira impeditiva para os gerentes fazerem uso de tais modelos. (SALTZMAN & MEHROTRA, 2004)

Para lidar com esse problema, os autores desenvolveram uma interface amigável no Excel para modelos de simulação (de eventos discretos) dinâmica. O modelo consiste em um sistema de filas genérico para o qual estão sempre disponíveis resultados analíticos; e a interface no Excel permite aos gerentes especificar interativamente uma grande variedade de parâmetros e analisar resultados, sem que eles fiquem expostos aos componentes do modelo de simulação.

Baseados em informações provenientes de gerentes operacionais de *call centers*, os autores foram capazes de utilizar esta estrutura para perguntar e responder algumas questões

empíricas importantes quanto ao comportamento operacional do sistema quando submetido a diferentes condições.

3.5.1.9. IVR Simulator Model

O **IVR Simulator Model** é um dos modelos de simulação desenvolvidos pela Database Systems Corporation para assistir seus clientes quando eles estão avaliando a compra de equipamentos para *call center*. Seu propósito é funcionar como uma espécie de calculadora de recursos. Ele calcula as exigências de recursos para uma campanha de chamadas, estimando o número de linhas requeridas para lidar com o processamento destas chamadas durante o horário de pico. O modelo assume que as chamadas são uniformemente distribuídas durante esse período. (DATABASE SYSTEMS CORP, 2005)

Segundo a fonte, para rodar as simulações, são necessárias certas informações – acuradas – a respeito dos clientes:

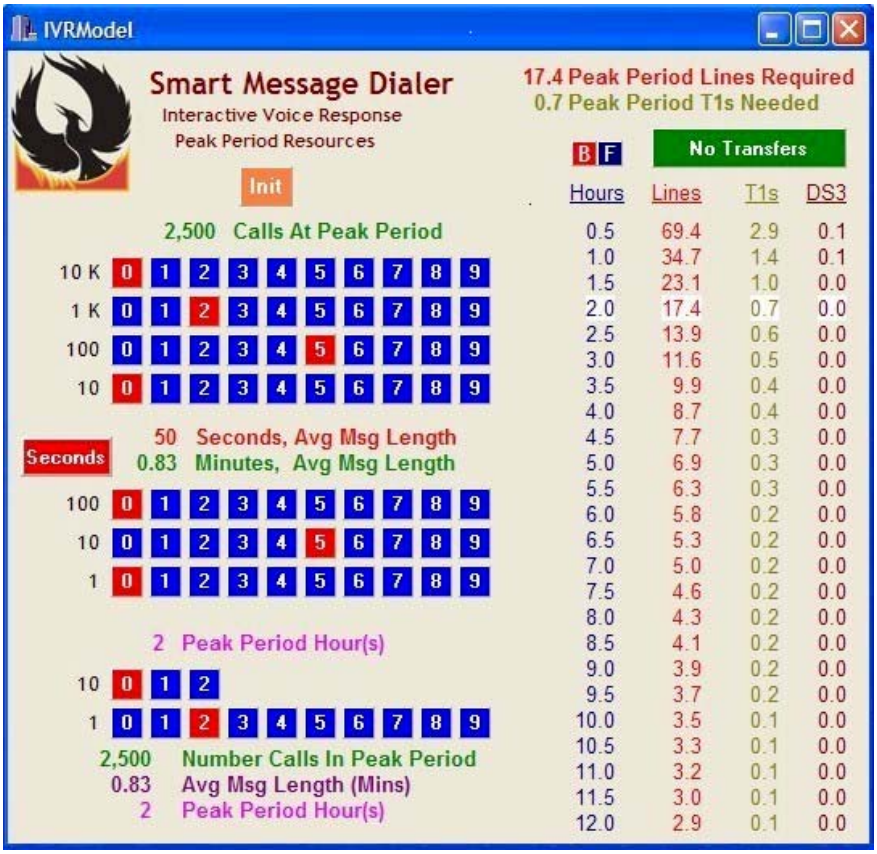
- duração média das chamadas
- número de chamadas recebidas durante o período de pico
- duração do período de pico

A simulação retorna – como pode ser observado na figura 9 a seguir – o número de linhas requeridas para suportar o volume de pico informado e também em outros dois cenários: um com um período de pico mais curto do que o previsto e outro com um período de pico mais longo do que o previsto. (DATABASE SYSTEMS CORP, 2005)

3.5.1.10. Quadros comparativos

Os quadros a serem apresentados têm por objetivo relacionar de forma sumarizada os atributos presentes em cada um dos *softwares* pesquisados e apresentados anteriormente nesta seção.

Figura 9 – Relatório de saída do IVR Simulator Model



Fonte: www.databasesystemscorp.com

O primeiro quadro – que pode ser visualizado na tabela 2 a seguir – apresenta as características gerais presentes (assinaladas com um "X") em cada *software*.

Tabela 2 – Características gerais dos *softwares* de simulação em *call centers* pesquisados

Software / Característica	Arena Contact Center	callLAB	SimACD	Service Model	Simul8	ccProphet	Contact Centers	EPSCC	IVR Simulator Model
Simplicidade		X				X		X	X
Foco em *Call Centers*	X	X	X			X	X	X	X
Terminologia de *Call Centers*	X	X	X			X	X	X	X
Preço acessível		X			X	X	X	X	X
Utilização em massa				X	X				
Flexibilidade (customização)							X	X	
Integração com outras ferramentas	X						X		
Modelagem em ambiente de planilha								X	
Modelagem gráfica	X			X	X	X			
Análise de *inputs*	X			X					
Indicadores de performance dinâmicos	X		X	X	X	X		X	
Animação gráfica	X			X	X	X			
Análise financeira	X		X	X					X

Fonte: Tabela elaborada pelo pesquisador

O segundo quadro – apresentado na tabela 3 a seguir – mostra os atributos da operação dos *call centers* que cada *software* é capaz de contemplar (assinalados com um "X").

Tabela 3 – Atributos operacionais contemplados pelos *softwares* de simulação em *call centers* pesquisados

Software / Atributo	Arena Contact Center	callLAB	SimACD	Service Model	Simul8	ccProphet	Contact Centers	EPSCC	IVR Simulator Model
Nível de serviço	X	X	X	X	X	X	X	X	X
Abandono	X	X	X	X	X	X	X	X	
Nível de utilização do sistema	X	X	X	X	X	X	X	X	X
Rotas (*scripts*)	X	X				X	X		
Transferência	X	X				X			
Tipos diferentes de chamadas	X			X	X	X			
Operação multi-site		X			X	X			
Skill dos agentes	X	X		X	X	X	X		
Skill-based routing	X	X			X	X			
Roteamento de rede		X							
"Re-tentativa" de ligação por parte do cliente	X	X	X						
Limitação na capacidade da central (clientes bloqueados)	X	X	X	X	X	X		X	X
Transbordo de ligações	X	X	X		X	X			
Operação *inbound* e *outbound*	X			X	X	X	X		
Políticas de discagem							X		
Mensagem deixada pelo cliente	X								
Escolha da distribuição do tempo de atendimento	X	X	X	X	X	X	X	X	
Escolha da distribuição do tempo entre chegadas	X	X		X	X	X	X	X	
Conferência (cliente e mais de um atendente)	X					X			

Fonte: Tabela elaborada pelo pesquisador

3.6. Resumo

A primeira seção deste referencial teórico – **Métodos de previsão de demanda** – apresentou alguns métodos genéricos de previsão de demanda existentes na literatura, incluindo séries temporais e séries causais. O início da seção trouxe também um breve histórico do assunto, a sua tipologia e destaca a sua importância, além de fornecer algumas dicas para o seu gerenciamento.

Foi possível constatar a importância de fazer previsões sobre eventos futuros, para que as projeções possam ser incorporadas ao processo de tomada de decisão. Dentre os principais métodos quantitativos de previsão, a regressão múltipla revelou-se bastante poderoso, sendo capaz de identificar os fatores que se relacionam com a variável a ser prevista.

Variáveis *dummy* podem ser utilizadas para registrar a ocorrência ou não de alguma característica. Os principais indicadores de qualidade do método são o F de significação, o R-quadrado e o valor p das variáveis explicativas. Um indicador capaz de mensurar a magnitude do erro de previsão incorrido pelo método é fundamental e um dos mais utilizados é o MAPE.

O objetivo da seção seguinte – **Previsão de demanda de *call centers*** – foi aplicar o assunto trabalhado na seção anterior na indústria de *call centers*. A importância desta aplicação é vital, já que consiste em um dos principais problemas atacados nesta pesquisa. Desta forma, foi discutida a utilidade do processo de previsão em *call centers* e, em seguida, foram apresentados os possíveis horizontes de tempo com que as previsões podem trabalhar, assim como as diferentes profundidades e níveis de sofisticação que os métodos existentes podem assumir.

O texto incluiu também dicas e alertas para algumas possíveis armadilhas e sugeriu um roteiro a ser seguido durante o processo de previsão. A seção ainda destacou a importância dos erros de previsão e algumas maneiras de mensurá-los. E finalizou com algumas aplicações reais, presentes na literatura, de métodos de previsão de demanda a *call centers*.

A relevância do processo de previsão de demanda (assim como a sua criticidade) ficou clara, já que esta consiste um dos mais importantes *inputs* para o dimensionamento da capacidade de atendimento. As maiores dificuldades acontecem por causa das grandes flutuações dos valores e da ocorrência de picos de demanda.

Outra constatação importante diz respeito ao fato dos métodos de séries temporais serem os mais usados nesta indústria, onde a demanda costuma ser prevista para cada bloco pequeno de tempo (15 ou 30 minutos). Um último alerta apontou para a necessidade de prever não somente a quantidade de ligações, mas também o tempo médio de atendimento.

A terceira seção – **Teoria de filas** – introduziu brevemente o histórico da abordagem e conceitualizou – de forma genérica – seus modelos analíticos, indicando os parâmetros envolvidos nos mesmos, assim como os seus *inputs* e *outputs*. A importância de tal assunto é justificada por se tratar de uma abordagem alternativa (em relação à simulação) no tratamento de alguns problemas quantitativos, como o de dimensionamento de capacidade.

Alguns modelos existentes na literatura foram citados, com variação dos seus parâmetros. A parte final tratou de aplicações reais da teoria de filas a problemas de gerenciamento de *call centers* e discutiu algumas questões envolvidas.

Foi importante descobrir que as premissas adotadas nas análises baseadas nos modelos analíticos de teoria de filas são extremamente limitadas quando trabalhadas no contexto atual dos *call centers* porque a maior parte dos modelos assume que as chegadas seguem uma distribuição exponencial, são todas do mesmo tipo, nunca abandonam a fila e são atendidas sempre pela ordem de chegada por agentes que atendem todas as chamadas da mesma maneira.

Ainda assim – e não obstante a verificação de que as disciplinas das filas, quando bem gerenciadas, tornam-se fortes aliadas da área de planejamento e controle da produção dos *call centers* – muitas empresas continuam apoiando as complexas tomadas de decisão acerca da alocação de recursos através dos modelos analíticos de teoria de filas, motivadas – principalmente – pela facilidade e rapidez da abordagem.

A seção de nome **Simulação** apresentou resumidamente o histórico da ferramenta, além da sua conceituação e tipologia. Algumas vantagens e desvantagens foram apontadas, bem como alguns exemplos de aplicação da ferramenta.

O texto permitiu constatar que a simulação é capaz de "comprimir o tempo" ao tentar cumprir o seu objetivo de melhorar o desempenho do sistema e de realizar análises de sensibilidade acerca dos dados de entrada, sem a necessidade de modificar efetivamente o sistema estudado (ou até mesmo antes de ele existir, na prática). Apesar de ser uma abordagem relativamente

cara e de difícil implementação, ela revelou-se mais adequada do que os métodos analíticos quando os processos a serem estudados são muito complexos e/ou apresentam altos níveis de incerteza.

A última seção – **Simulação em *call centers*** – tratou especificamente da aplicação da ferramenta de simulação aos problemas de dimensionamento e gerenciamento dos *call centers*. Desta forma, um retrato da atual importância da aplicação foi mostrado e as causas que levaram a esse importante *status* foram levantadas, assim como foram especuladas as tendências futuras do tema.

Foi também discutido o grande potencial que a ferramenta apresenta em relação ao gerenciamento dos problemas da indústria em questão, incluindo os principais aspectos dos *call centers* que a ferramenta é capaz de levar em consideração. Uma comparação entre as possibilidades da ferramenta e as dos modelos analíticos foi apresentada, dentro do escopo das centrais de atendimento.

A última parte da seção resumiu uma extensa pesquisa acerca dos *softwares* de simulação para *call centers* existentes no mercado, destacando e comparando sua funcionalidade e adequabilidade aos problemas da indústria em questão, levando em conta também as características operacionais das centrais de atendimento que cada um é capaz de contemplar.

A seção permitiu perceber que a utilização da simulação para auxiliar a tomada de decisões no gerenciamento de *call centers* possibilita, entre outras: a avaliação da situação atual; a análise de cenários; a previsão da necessidade agregada de recursos e a programação da força de trabalho; e a mensuração de indicadores de performance.

Também foi possível verificar que um modelo de simulação pode contemplar alguns aspectos críticos dos *call centers* modernos, como: nível de serviço específico; flexibilidade na distribuição dos tempos entre chegadas e de atendimento; tipos múltiplos de chamadas; padrões de abandono de chamadas e de retorno de ligações; e *overflow* da central.

Por isso e por causa de outras tendências recentes e, de uma forma geral, da sofisticação e complexidade cada vez mais presentes nos sistemas de *call centers*, os métodos tradicionais mais usados no seu gerenciamento estão se tornando obsoletos e perdendo espaço para a simulação, principalmente quando a operação é detalhada, a demanda varia muito e é

necessário viabilizar um melhor entendimento acerca dos processos da central e gerar resultados mais acurados a respeito da sua performance.

Em relação aos *softwares*, foi constatado que pacotes de simulação especificamente desenvolvidos para *call centers* estão disponíveis no mercado, permitindo analistas pouco experientes construir modelos de centrais de atendimento, com relativa facilidade.

A comparação entre as opções disponíveis revelou, entre outras coisas, que a maioria delas empreende a terminologia própria de *call centers*, mas não é utilizada maciçamente por usuários. Além disso, quase nenhuma opção se mostrou flexível e capaz de ser integrada a outras ferramentas de gerenciamento. A modelagem e a animação gráficas estão disponíveis em apenas algumas opções. Os indicadores de performance podem ser observados dinamicamente na maioria dos *softwares*, mas apenas alguns são capazes de realizar análises financeiras (de custos) e quase nenhum possui alguma ferramenta para fazer um tratamento estatístico preliminar dos dados de entrada.

Adicionalmente, a comparação permitiu observar que quase todos os *softwares* contemplam o comportamento de abandono por parte do cliente. A maioria também permite a escolha da distribuição dos tempos de atendimento e entre chegadas, assim como são capazes de limitar a capacidade da central e determinar que alguns clientes sejam bloqueados antes de entrar no sistema. Algumas das opções consideram diferentes tipos de centrais (*inbound* e *outbound*) e de chamadas e habilidades dos agentes, permitindo o roteamento das ligações (inclusive o baseado em habilidades), que podem ser transbordadas. Poucos *softwares*, no entanto, lidam com centrais em diferentes locais físicos e permitem que diferentes políticas de discagem sejam testadas (para os *call centers* do tipo *outbound*). Poucos também são aqueles capazes de contemplar algumas características das centrais, como: a "re-tentativa" de ligação por parte do cliente (depois de ser bloqueado ou desistir de esperar), mensagens deixadas pelos clientes, conferência (entre o cliente e mais de 1 atendente) e transferência de chamadas.

Entre os *softwares*, o Arena Contact Center foi o que conseguiu reunir mais atributos presentes nas operações das centrais de atendimento, mas deixou a desejar em alguns aspectos gerais, como simplicidade, flexibilidade e utilização massificada.

4. Metodologia

4.1. Tipo de pesquisa

Vergara (1997) sugere dois critérios segundo os quais as pesquisas na área de administração podem ser classificadas: quanto aos fins e quanto aos meios.

De acordo com a mesma, uma pesquisa, de uma forma geral, pode ter sua finalidade definida como: exploratória, descritiva, explicativa, metodológica, aplicada e/ou intervencionista.

Segundo a autora, uma pesquisa será explicativa quando tiver como objetivo justificar os motivos de algum fenômeno, esclarecendo quais fatores contribuem para a sua ocorrência; será metodológica quando se referir a instrumentos de captação ou manipulação da realidade; e será aplicada quando for motivada pela necessidade de resolver problemas concretos, práticos.

Assim, o presente estudo pode ser caracterizado, do ponto de vista da sua finalidade, como: **explicativo**, porque busca justificar o uso de métodos específicos de previsão de demanda e da ferramenta de simulação no ambiente de *call centers* e entender que fatores e características da operação contribuem (ou não) para isso; **aplicado**, pois foi motivado pela necessidade de resolver problemas práticos na operação de *call centers*: a previsão de demanda e o dimensionamento da capacidade de atendimento; e **metodológico**, já que elabora e propõe a implementação de duas ferramentas – de previsão e de simulação – que consistem em instrumentos de captação (previsão) e manipulação (simulação) da realidade.

A primeira ferramenta busca reunir informações quantitativas tratáveis (atualmente consideradas apenas de forma subjetiva) acerca de uma das operações da empresa e processá-las metodicamente de forma a conseguir prever de forma mais acurada a demanda de ligações para a central, conforme será detalhado na seção 6.1 a seguir.

A segunda objetiviza reproduzir o funcionamento de uma outra operação da empresa em alguns horários e sob algumas condições e simular cenários alternativos quanto à oferta de capacidade de atendimento proporcionada, para verificar a performance do sistema perante os diferentes cenários de forma a poder escolher o dimensionamento ideal para a central, conforme será abordado na seção 6.2 a seguir.

De acordo com Vergara (1997), uma pesquisa pode ser classificada – quanto aos meios de investigação – como sendo: pesquisa de campo, pesquisa de laboratório, pesquisa telematizada, investigação documental, pesquisa bibliográfica, pesquisa experimental, investigação *ex post facto*, pesquisa participante, pesquisa-ação e/ou estudo de caso.

Ainda segundo a autora, uma pesquisa será de campo quando consistir em uma investigação empírica realizada no local onde ocorre um fenômeno, incluindo, por exemplo, entrevistas, questionários, observação participante; e será um estudo de caso, quando o estudo estiver circunscrito a uma ou poucas unidades (pessoas, produtos, empresas), e for aprofundado e detalhado.

A presente pesquisa, então, pode ser entendida – no que diz respeito aos meios – como sendo: **de campo**, já que é baseada em entrevistas e em uma investigação empírica no local do fenômeno; e também como **estudo de um caso único**, já que está restrito a apenas uma unidade (empresa), permitindo uma maior profundidade e um maior nível de detalhes durante a análise.

Yin (1994) defende que o caminho ideal para identificar a ferramenta mais adequada para um estudo é analisar três critérios de pesquisa: o tipo de questão de pesquisa; o grau de controle que o pesquisador tem sobre o comportamento dos eventos; e o foco da pesquisa em eventos históricos ou contemporâneos.

O presente trabalho envolve perguntas do tipo "como" e "por que", caracterizando-se como explicativo; também oferece baixo grau de controle sobre os eventos (decisões gerenciais); e se propõe a contemplar e analisar situações restritas a um período específico de tempo (eventos contemporâneos). O estudo de caso explanatório (ou explicativo) é a ferramenta ideal para um trabalho com essas características, sendo considerado uma metodologia abrangente, que vai desde a pergunta de pesquisa até as conclusões e a redação final, passando pela coleta e análise de dados (YIN, 1994).

4.2. Seleção do caso e dos sujeitos

Yin (1994) coloca que os estudos de casos podem limitar-se a uma ou a várias unidades de análise. Segundo o autor, o caso único deve ser utilizado para fins explicativos, quando o estudo se enquadrar em pelo menos uma dentre três opções: caso decisivo, caso extremo ou

caso revelador. O primeiro tipo ocorre quando há elementos específicos que capacitam a pesquisa como instrumento de validação ou contestação da teoria em questão.

Um dos objetivos do presente estudo é verificar a adequabilidade da ferramenta de simulação à atividade de dimensionamento da capacidade operacional de uma central de atendimento telefônico. A idéia consiste em utilizar o caso como instrumento de validação ou contestação dessa adequabilidade, de forma a verificar que aspectos operacionais do *call center* são determinantes para justificar o uso da simulação, o que o caracteriza como decisivo. Como a pesquisa também apresenta uma finalidade explicativa (conforme abordado anteriormente na seção 4.1), a escolha de apenas uma unidade de análise parece a mais apropriada, até para que seja garantida uma profundidade necessária para os seus propósitos.

A unidade de análise abordada na presente pesquisa é a matriz da Contax Contact Center (do grupo Telemar), a maior empresa fornecedora de serviço de atendimento telefônico do Brasil. Alguns motivos guiaram a escolha da (única) unidade de análise:

- representatividade – pela sua ampla abrangência no mercado nacional, a Contax pôde ser escolhida como unidade única de análise, mantendo mesmo assim um bom nível de representatividade dentro da indústria;

- heterogeneidade operacional – por causa da grande variedade da sua carteira de clientes, a Contax trabalha com diferentes operações, com níveis de complexidade bem distintos entre si, fato que permitiu testar as ferramentas em cenários com características diferentes;

- acessibilidade – pelo fato de alguns ex-alunos do COPPEAD trabalharem atualmente na Contax, o acesso às pessoas responsáveis pela previsão de demanda e pelo dimensionamento e gerenciamento da capacidade de atendimento foi facilitado.

Os principais sujeitos do caso são:

- o sr. Alberto Porto, Gerente de Planejamento de Tráfego e responsável pela previsão de demanda e dimensionamento da capacidade de atendimento;

- e alguns de seus coordenadores (também responsáveis pela previsão de demanda e planejamento de capacidade, mas cada qual com o seu horizonte de planejamento – curto ou longo prazo – abrangência geográfica e carteira de produtos) e analistas:
    - o sr. Alessandro Meira, Coordenador de Planejamento de Longo Prazo – Telemar e OI;
    - a sra. Vânia Miranda, Coordenadora de Planejamento de Curto Prazo – Sudeste;
    - o sr. Mario Pires, Analista de Planejamento de Longo Prazo – 102;
    - e a sra. Renata Lourenço, Analista de Planejamento de Curto Prazo – Sudeste.

4.3. Coleta e tratamento dos dados

Nas pesquisas de campo, segundo Vergara (1997), os dados podem ser coletados através de: observação (simples ou participante), questionário, entrevista ou formulário. Na modalidade de entrevista, a presença física simultânea do autor e do(s) entrevistado(s) é obrigatória. Ela pode acontecer sob a forma informal, focalizada (em um assunto) ou por pautas (estruturada, seguindo uma agenda, e obtendo maior profundidade).

Em relação à coleta de dados para o estudo de caso em questão, foi escolhido o modelo de entrevista (pela possibilidade de presença física simultânea) focalizada (pela necessidade de uma certa profundidade nos assuntos de interesse da pesquisa). Os sujeitos mencionados na seção 4.2 foram entrevistados (com e sem gravação) em 4 oportunidades, entre o período de maio de 2005 e agosto de 2006.

É descrita a seguir a maneira como as informações obtidas foram tratadas de forma a responder as questões levantadas. Aqui, essas questões também estão divididas em grupos, conforme foi feito na seção 1.1.2:

4.3.1. Previsão

a) Que metodologia(s) é(são) aplicada(s)?

Para responder a primeira questão, o gerente e os profissionais responsáveis pela previsão de demanda foram argüidos a respeito.

b) Essa(s) é(são) a melhor forma de prever tais grandezas?

c) Se não for(em), qual é?

Em função da resposta obtida na primeira questão, do conhecimento detalhado (através de observação do seu histórico) das características dessas grandezas e dos resultados de uma pesquisa relacionando as melhores técnicas de previsão de demanda de grandezas com as características levantadas, foi respondida a segunda questão, assim como a terceira, junto com a qual foi desenvolvido e sugerido um modelo de previsão para a quantidade de chamadas demandadas (e também para o tempo médio de atendimento) por unidade de tempo.

4.3.2. Dimensionamento

d) Qual metodologia é aplicada para a resolução desse problema?

e) A simulação é utilizada?

f) Se não for, por que não é?

Para responder as três primeiras questões, o gerente e os profissionais responsáveis pelo dimensionamento da capacidade de atendimento foram argüidos a respeito.

g) Quais são os *trade-offs* envolvidos na decisão de qual metodologia usar (simulação *versus* outras)?

Para responder essa questão, foi realizada uma pesquisa relacionando as principais vantagens e desvantagens da simulação em comparação ao uso de abordagens analíticas no contexto de *call centers*, contemplando as peculiaridades deste setor.

h) Como a simulação pode dar apoio às decisões referentes ao processo de dimensionamento do *call center*?

Para responder essa questão, diversos modelos de simulação foram construídos – contemplando diferentes características reais dos *call centers* e para diferentes possíveis cenários alternativos – para auxiliar a tomada de decisão a respeito do dimensionamento do

efetivo. Eles foram usados para calcular indicadores de performance e sugerir soluções para o problema em questão, em relação a situações efetivamente vividas pela empresa no passado recente.

i) Que aspectos operacionais do *call center* não podem ser bem tratados por outras metodologias, sendo determinantes para justificar o uso da simulação?

j) A simulação fornece resultados mais acurados em algumas situações do cotidiano operacional do *call center*? Em quais situações?

Para responder essas questões, os mesmos indicadores e resultados obtidos a partir das soluções propostas pela abordagem analítica (utilizada atualmente pela empresa) foram comparados aos conseguidos através da simulação e aos resultados que, de fato, ocorreram posteriormente na operação da empresa, em cenários bem parecidos com os usados nas análises. Essa confrontação com os resultados reais foi possível graças a uma incursão na base de dados da Contax, que permitiu o acesso às informações relevantes.

A acurácia das duas abordagens (analítica *versus* simulação) foi, então, comparada em uma série de cenários (idealizados a partir da inspiração adquirida durante o desenvolvimento da pesquisa bibliográfica apresentada na seção 3.5 anterior) envolvendo diferentes características – reais e supostas – da operação e situações hipotéticas de comportamento da demanda, para que fosse feito um mapeamento a respeito de quais características das operações de *call centers* e quais possíveis cenários hipotéticos estariam sendo responsáveis por garantir mais acurácia para os resultados encontrados e por justificar o uso da abordagem experimental.

k) Quais *softwares* estão disponíveis para o uso da simulação aplicada a *call centers*?

l) Quais são as suas características?

m) Quais deles são utilizados no *call center* estudado?

Para responder a última questão, os responsáveis pelo setor de dimensionamento foram argüidos a respeito. Uma pesquisa (apresentada na seção 3.5.1) a respeito dos *softwares* existentes ajudaram a responder as outras duas, sendo que a resposta à questão "l" ficou mais evidenciada nos quadros apresentados nas tabelas 2 e 3 anteriores.

O *software* escolhido para modelar e simular as operações de atendimento telefônico da Contax foi o Arena Contact Center. Tal escolha deveu-se à ótima adequação do produto às especificidades das operações em questão, à facilidade do seu manuseio, à sua riqueza de recursos (conforme pode ser observado nas tabelas 2 e 3 anteriores), ao prévio conhecimento do pesquisador a respeito do *software* Arena (que serve de plataforma para o produto) e ao pronto fornecimento de uma cópia de estudante por parte da empresa fornecedora do *software*, o que efetivamente viabilizou a pesquisa.

4.4. Limitações do método

Uma das críticas mais comuns ao método do caso trata da impossibilidade de generalização dos resultados da pesquisa. Adicionalmente, o estudo de um único caso carrega a grave limitação da incerteza a respeito do fato do caso estudado ser realmente representativo do universo do qual ele seria uma amostra. (GIL, 1987)

A presente pesquisa foi conduzida em apenas uma empresa e pode, portanto, não ser totalmente representativa em termos de cenário nacional da indústria de *call centers*. Pelo fato de se tratar da maior empresa brasileira no ramo e de apresentar uma ampla variedade de clientes (tanto em termos de tamanho quanto em termos de natureza da sua operação), a probabilidade de não representatividade, no entanto, diminui razoavelmente.

Outras limitações decorrentes do escopo restrito do estudo podem ser explicitadas. Em primeiro lugar, algumas das conclusões obtidas a partir das entrevistas dependem em boa parte das opiniões pessoais dos indivíduos que foram argüidos na ocasião da pesquisa, estando sujeitas a uma boa dose de subjetividade. Adicionalmente, algumas constatações podem carregar um caráter um tanto quanto temporal, já que a mesma empresa pode vir a apresentar um cenário completamente diferente em um outro momento de tempo em que esse texto venha a ser lido, que não aquele a partir do qual ele foi elaborado.

A ferramenta de previsão proposta pela pesquisa considera apenas informações referentes às características intrínsecas das datas consideradas na análise (como dia da semana, feriados e proximidade com os dias de chegada e vencimento das contas telefônicas). Outros eventos especiais relacionados a datas específicas, mas não com um caráter intrínseco – como campanhas na mídia, crescimento de base e outros aspectos potencialmente impactantes na

demanda – ainda precisam ser considerados "por fora" pelos analistas e inseridos manualmente na ferramenta.

A ferramenta em questão tratou os dias considerados na análise de forma binária (sim ou não) no tocante a consistirem ou não de feriados; ou seja, não foi considerado um meio-termo para essa questão, impossibilitando a ocorrência de "feriados menos intensos", o que normalmente pode ocorrer com feriados locais ou com dias "imprensados" entre o feriado e o final de semana.

O efeito da proximidade de certos dias em relação às datas de chegada e vencimento das contas de telefone foi considerado apenas de forma nacional pela ferramenta. Em outras palavras, o impacto na demanda proporcionado pelo fato de um determinado dia tratar-se da véspera do vencimento para 100.000 contas de Minas Gerais pode ser diferente do impacto causado pelo fato do dia tratar-se da véspera do vencimento para 100.000 contas do Ceará, já que o comportamento cultural dos mineiros em relação ao vencimento da conta pode ser diferente do comportamento dos cearenses; mas a ferramenta não reconhece essas potenciais diferenças e trata os impactos de forma homogênea.

Devido à dificuldade de obtenção – junto à Contax – de dados mais recentes do histórico de demanda (volume de ligações e TMA) do produto 103, não foi possível verificar a qualidade do modelo de previsão através de dados que não foram usados para a construção do mesmo. Só foi possível mensurar sua funcionalidade em relação ao passado (período dos dados usados como *input*), mas não se o modelo vai funcionar bem em relação ao futuro (o que teria sido conseguido se o modelo pudesse ter sido aplicado a um período de dados diferente dos que foram usados para a sua concepção). Em outras palavras, o modelo pode ter se adequado bem a um conjunto específico de dados (datas), mas não foi possível generalizar tal adequabilidade.

Ainda a respeito da ferramenta de previsão proposta, não é razoável deixar de atentar para o fato que ela não considera uma possível sazonalidade da demanda ao longo do ano. Pelo fato do histórico de demanda disponível para a sua elaboração consistir de apenas alguns poucos meses, não foi possível para a ferramenta tentar capturar esse efeito sazonal. Mas não é sensato afirmar que ele não existe.

Já em relação ao problema de dimensionamento, a cópia de estudante conseguida junto à empresa fornecedora do *software* usado para a simulação das operações (Arena Contact Center) permite apenas a modelagem de centrais de atendimento com, no máximo, 20 posições de atendimento (PAs). Esta limitação impediu a simulação da operação em horários de muito movimento, que requereriam uma quantidade maior de operadores.

Isto também impossibilitou a simulação de um dia contínuo de operação da central (que consistia na idéia original), já que em alguns períodos do dia, seriam dimensionados muito mais do que 20 atendentes. Tal impedimento foi um pouco lamentável, porque a simulação do dia, como um todo, permitiria a visualização das transições de horários, o que seria bastante interessante por causa das mudanças acentuadas nas curvas de demanda experimentadas durante as transições, que permitiriam enriquecer a análise.

Esta restrição, no entanto, não chegou a ser comprometedora, pois o interesse maior da pesquisa consiste na análise de situações com diferentes níveis de atendimento à demanda (e conseqüentes níveis de serviço oferecidos); assim – já que não era necessário modelar, necessariamente, situações com efetivos substancialmente diferentes – a limitação não impediria o dimensionamento de, por exemplo, 20 agentes para um horário que necessitava de 10 deles, novamente 20 para um horário que necessitava de 20 deles e outra vez 20 para um horário que necessitava de 40 deles.

Durante a simulação dos modelos, não foi possível descobrir a correta distribuição estatística e a variabilidade do tempo entre chegadas de ligações e do tempo de atendimento. Isso aconteceu porque a Contax não dispunha – ao menos de forma acessível – do registro de cada ligação individual, contendo a informação do horário de sua entrada no sistema e da quantidade de tempo transcorrida até o seu atendimento.

Estas informações estavam disponíveis apenas por agrupamentos de 30 minutos (por exemplo: 792 ligações chegaram ao sistema no período entre 10:00 e 10:30 de um determinado dia, levando, em média 28,7 segundos para serem atendidas), o que impossibilitou uma análise de aderência destes tempos aos formatos conhecidos de distribuições.

Para lidar com esta limitação, os modelos assumiram a premissa comumente utilizada de que o intervalo entre chamadas segue uma distribuição exponencial e o tempo de atendimento segue uma distribuição Erlang.

No primeiro caso, a suposição é bem razoável já que o processo de chegada das ligações não apresenta nenhum controle e pode ser considerado aleatório. Mas, no segundo caso, apesar de a distribuição Erlang representar bem o processo de atendimento, a suposição é um pouco mais forte e ousada. Além disso, tal distribuição requer um parâmetro adicional (além da média) a respeito da variabilidade dos dados, que teve que ser suposto, devido à falta de registro das ligações individuais, conforme mencionado anteriormente.

O ideal seria que este registro pudesse ser utilizado para a correta investigação acerca do formato das distribuições destes tempos e dos seus respectivos parâmetros, até porque – conforme levantado durante a pesquisa bibliográfica – os resultados obtidos através da simulação podem ser sensíveis a estes formatos e, portanto, a acurácia das conclusões pode ter ficado um pouco prejudicada.

Durante a mesma atividade, foi necessário adotar premissas baseadas no julgamento subjetivo dos analistas (e, conseqüentemente, um tanto arbitrárias) a respeito do comportamento de retorno da ligação por parte do cliente após o abandono da chamada (depois de aguardar por muito tempo) e também após a chamada ser bloqueada (antes mesmo de entrar no sistema) por causa da ausência de linhas disponíveis no tronco.

O sistema da Contax não permite o rastreamento individual de cada chamada e, portanto, não foi possível mapear objetivamente estes comportamentos. A sua correta consideração, no entanto, poderia gerar resultados mais acurados.

O julgamento subjetivo dos analistas também precisou ser utilizado para estimar o impacto no tempo de atendimento ocorrido quando as chamadas não são atendidas pelo seu tipo de agente preferencial ("básico" ou "*plus*"). O procedimento ideal seria uma coleta empírica destas informações, dificultada, no entanto, pelo fato dos atendentes estarem situados em uma cidade diferente da onde se encontravam a equipe de dimensionamento e o pesquisador.

5. Descrição do Caso

Este capítulo descreve – de uma forma geral e enfocando os aspectos mais relevantes para os objetivos da pesquisa – a empresa estudada, a partir da coleta de informações realizada principalmente através de entrevistas e pesquisas na Internet, mas incluindo também algumas observações realizadas pelo pesquisador.

5.1. A Gerência de Planejamento de Tráfego da Contax

A área da empresa enfocada e explorada neste estudo é a de Planejamento de Tráfego, já que é responsável pelo tratamento dos problemas aqui explorados, em especial o de previsão de demanda de chamadas e o de planejamento da capacidade de atendimento. O principal *software* que o setor utiliza para auxiliar as decisões gerenciais de planejamento operacional é o TotalView, totalmente integrado à infra-estrutura de telefonia.

O produto consiste em uma ferramenta de força de trabalho utilizada para dimensionar e controlar o atendimento, que auxilia os gerentes na previsão de demanda (por horário e dia da semana), no planejamento da capacidade e na análise de cenários (através de fórmulas Erlang), no desenvolvimento do esquema de horários dos atendentes, na integração entre diferentes centrais e com a geração de relatórios gerenciais. O *software* desenvolve seus cálculos matemáticos baseando-se em métodos analíticos (teoria de filas), não trabalhando com simulação. (IEX, 2005)

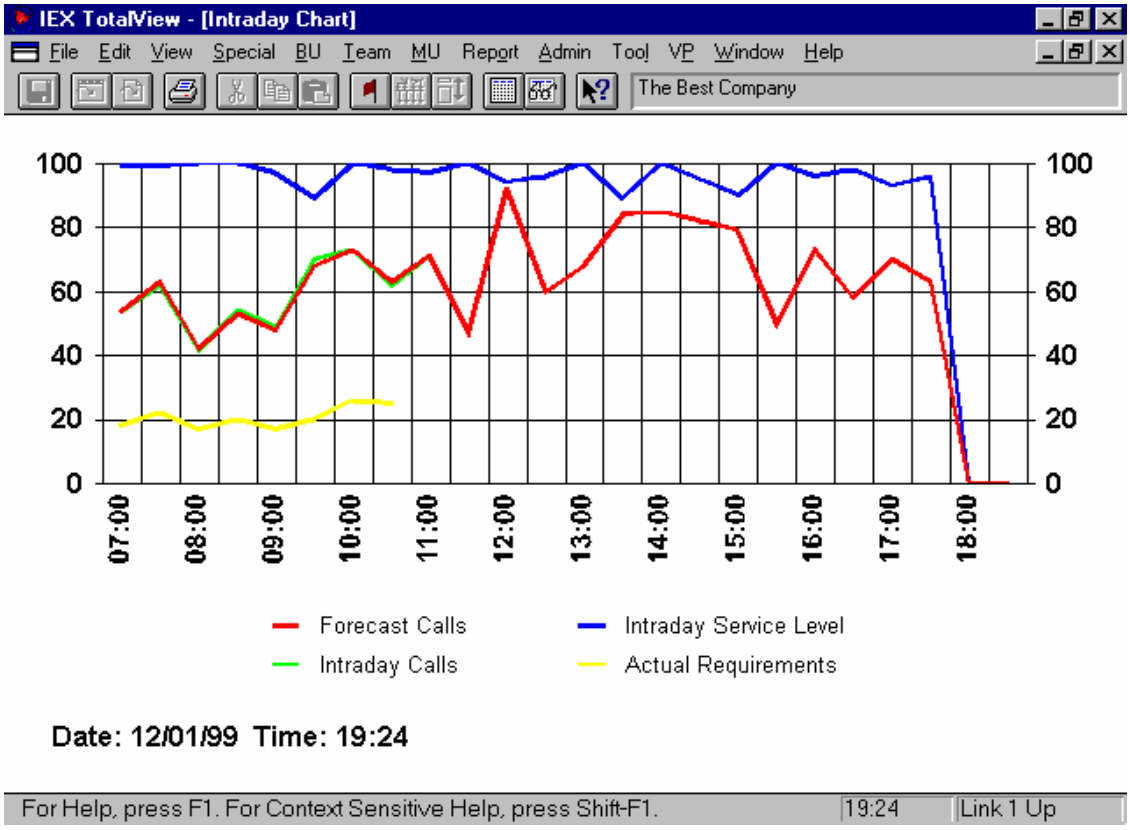
A arquitetura do TotalView está esquematizada na figura 10 a seguir.

Figura 10 – Arquitetura do TotalView

Fonte: Material coletado nas entrevistas

O *software* permite o acompanhamento da quantidade de chamadas, permitindo a comparação do que está efetivamente acontecendo com o que foi previsto e com o que foi acordado no nível de serviço, conforme pode ser observado na tela exemplifica na figura 11 a seguir.

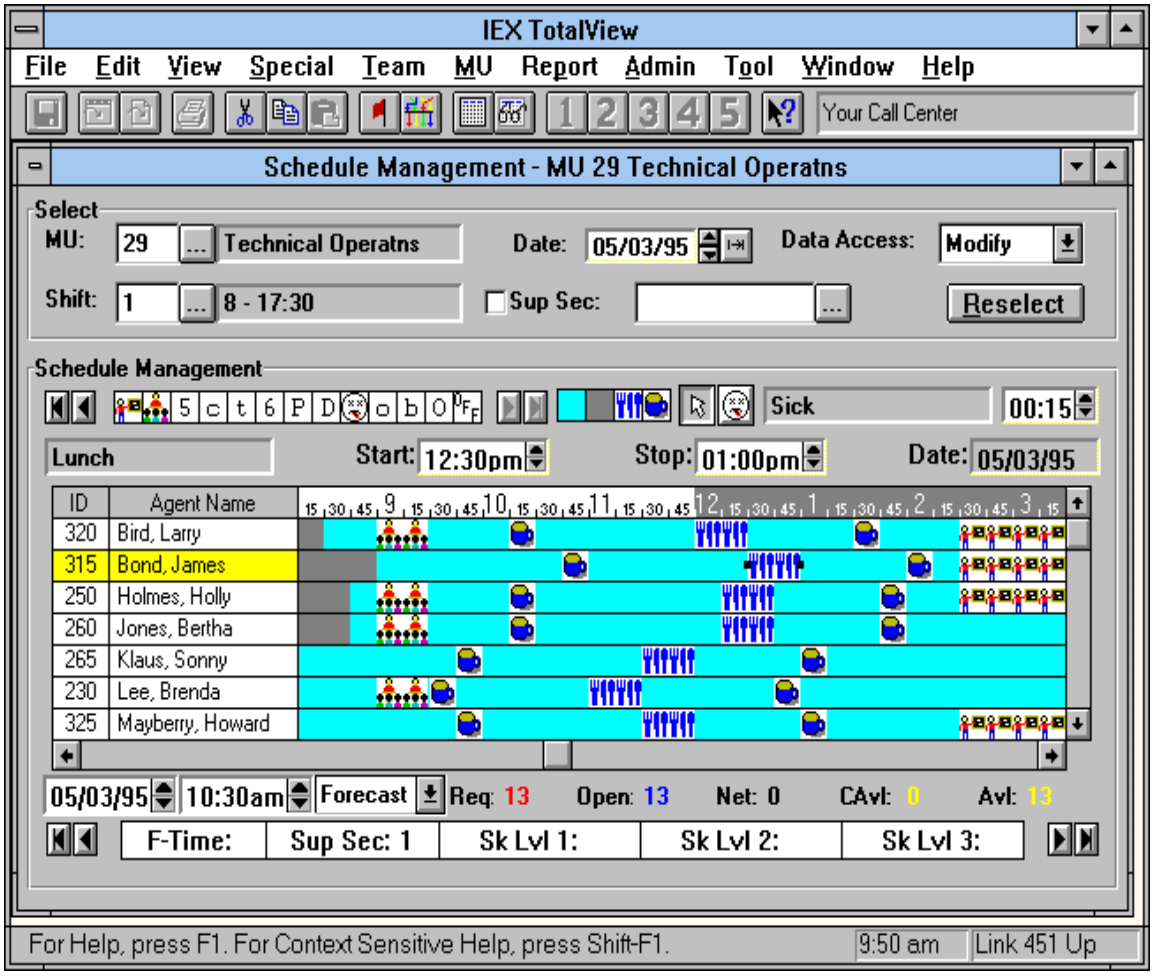
Figura 11 – Tela de acompanhamento da quantidade de chamadas do TotalView



Fonte: Material coletado nas entrevistas

Também é possível verificar a aderência, que consiste na comparação entre a escala planejada para cada atendente e o seu cumprimento efetivo de horários e tarefas, conforme pode ser observado na tela exemplificada na figura 12 a seguir.

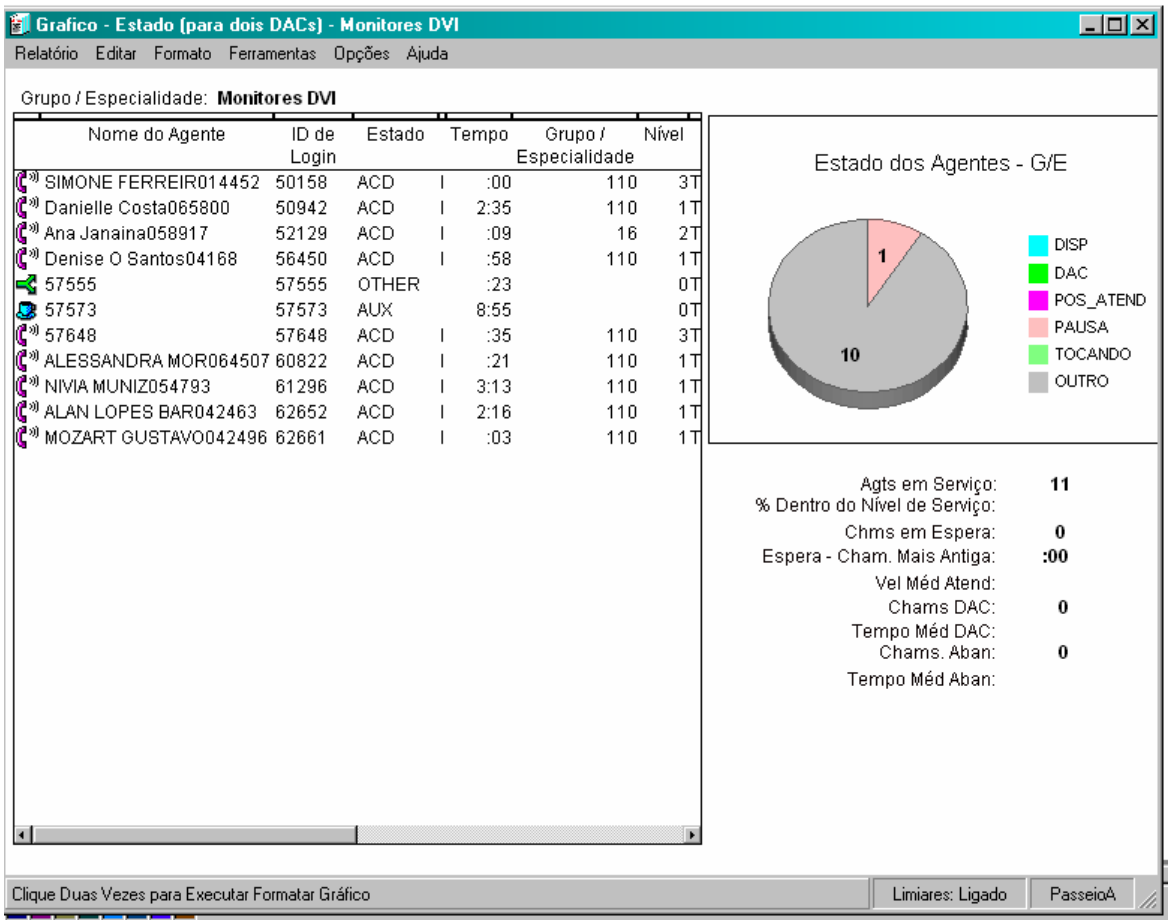
Figura 12 – Tela de verificação da aderência do TotalView



Fonte: Material coletado nas entrevistas

Outro *software* utilizado pela área de Planejamento de Tráfego é o CentreVu Supervisor, que monitora as operações e a coleta de dados (que podem ser visualizados na figura 13 a seguir) da central.

Figura 13 – Tela de monitoramento do CentreVu Supervisor



Fonte: Material coletado nas entrevistas

Estes dados são então organizados em relatórios que ajudam os analistas no gerenciamento dos recursos e do pessoal das centrais de atendimento. Os relatórios podem ser exibidos em tempo real (conforme exemplificado na figura 14 a seguir), impressos imediatamente ou armazenados.

Figura 14 – Relatório do CentreVu Supervisor

Resumo das Atividades

Relatório Editar Formato Ferramentas Opções Ajuda

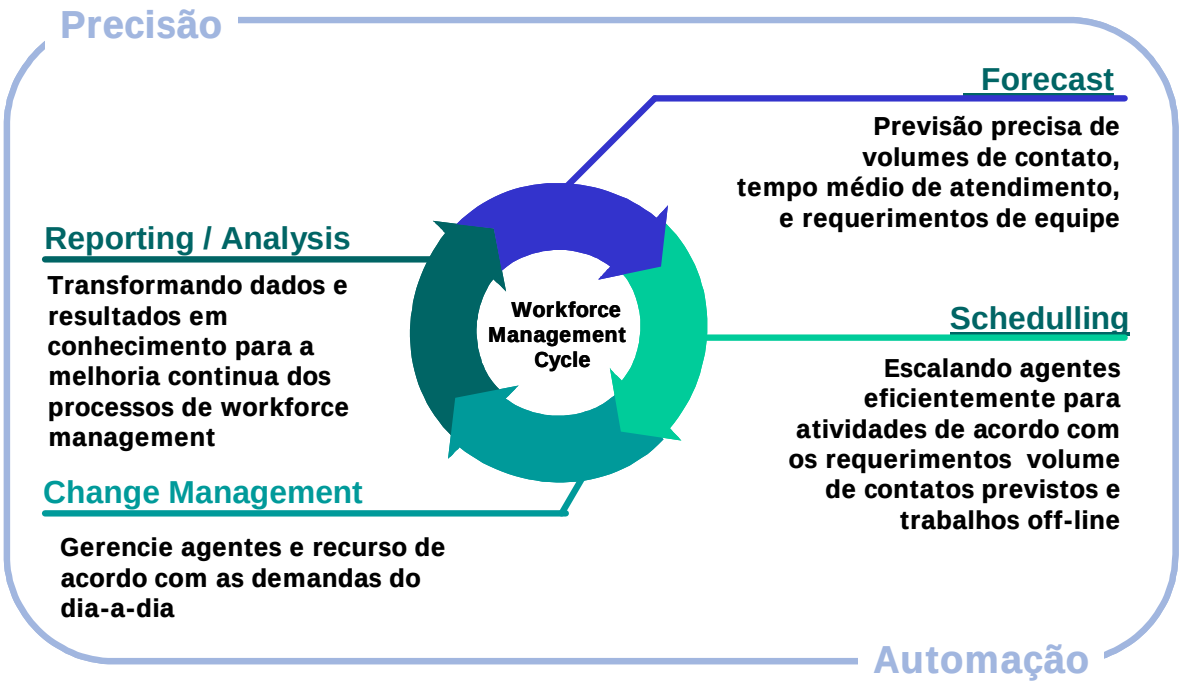
Grupo / Especialidade	98	100	101	106	107	115	119	136
Est. da Esp.	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL
Chms em Espera	0	0	0	0	0	0	0	
Cham Mais Antiga - Esperando	:00	:00	:00	:00	:00	:00	:00	:
V.Média Atend								
Chams DAC	0	0	0	0	0	0	0	
Tempo Médio DAC								
Chams. Aban	0	0	0	0	0	0	0	
Tempo Méd p/ Aban								
Agts Disponíveis	0	0	6	0	1	0	8	
Agts TOCANDO	0	0	0	0	0	0	0	
Agts em Chams DAC	0	0	0	0	0	0	0	
Agts em POS_AT	0	0	0	0	0	0	0	
Agentes em Outro	0	0	0	0	0	0	1	
Agentes em PAUSA	0	0	0	0	0	0	0	
Agts em Serv.	0	0	6	0	1	0	9	

Limiares: Ligado Telemar2

Fonte: Material coletado nas entrevistas

O principal objetivo da Gerência de Planejamento de Tráfego é ter a quantidade certa de pessoas, na hora certa e no lugar certo, dentro de uma previsão precisa, para atingir as metas de atendimento estabelecidas. Em outras palavras, o *call center* deve ser otimizado em termos de redução de custos e de atendimento dos clientes finais dentro do nível de serviço estipulado pelo contratante. Para atingir tais objetivos, as atividades do setor precisam ser encaradas internamente como parte de um processo de melhoramento contínuo, conforme ilustrado na figura 15 a seguir.

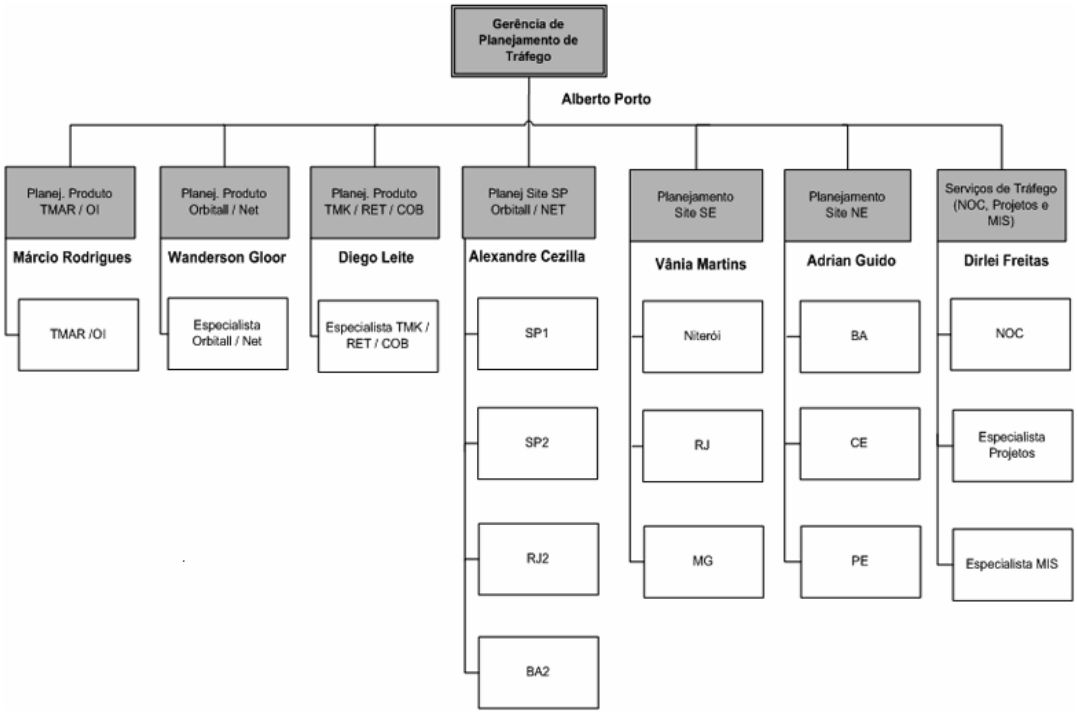
Figura 15 – Atividades da Gerência de Planejamento de Tráfego



Fonte: Material coletado nas entrevistas

A estrutura organizacional da área pode ser visualizada na figura 16 a seguir:

Figura 16 – Organograma da Gerência de Planejamento da Contax



Fonte: Material coletado nas entrevistas

Conforme pode ser observado, a gerência de planejamento central chefia várias coordenações que incluem, além do coordenador, analistas e especialistas:

- as três primeiras (da esquerda para a direita), de planejamento de longo prazo, responsáveis cada qual por uma fatia da carteira de produtos da empresa: Telemar & OI; Orbital & Net; e TMK, RET & COB, respectivamente;

- as três seguintes, de planejamento de curto prazo, responsáveis cada qual por uma parcela geográfica do mercado da empresa: São Paulo; Região Sudeste; e Região Nordeste, respectivamente;

- a última, responsável por serviços e projetos especiais da empresa.

As coordenações de planejamento de longo prazo realizam toda a parte de macro-previsão e macro-dimensionamento das operações dos seus respectivos produtos. O olhar dos coordenadores e analistas é estratégico, já que eles devem planejar o *call center* a médio (60 dias de antecedência) e longo prazo, e determinar antecipadamente quantas posições de atendimento, operadores, linhas e portas de URA a central irá demandar em um período futuro, tendo como base a captação de clientes e o crescimento natural da central.

Já as coordenações de planejamento de curto prazo trabalham mais o dia-a-dia, tendo a incumbência de aplicar o que foi planejado pelas coordenações de longo prazo (em termos de quantidade de recursos) e de identificar possíveis desvios em relação ao que foi planejado. Além disso, devem gerenciar aspectos cotidianos da operação (como horas-extras, folgas etc.). O olhar dos coordenadores e analistas é mais operacional, já que eles devem acompanhar a demanda de ligações que o *call center* recebe a fim de adequar os recursos humanos e tecnológicos a ela, sem sobras ou faltas.

Esta estrutura organizacional foi implantada no início de 2005. Anteriormente, não havia equipe de planejamento de longo prazo. As pessoas que faziam as previsões e dimensionamentos de longo prazo eram as mesmas que cuidavam do planejamento local, de curto prazo, cuidando do dia-dia, fazendo as escalas, programando horas-extras e acompanhando as filas. Assim, no momento em que havia chamadas em fila de espera, o planejamento de médio e longo prazo perdia prioridade e o foco acabava sendo no curto prazo. O "emergencial" sempre substituía o "importante", que era preterido. Dessa forma, as previsões eram piores, os desvios eram maiores e o que era "importante" (mas havia sido

deixado de lado) virava “emergencial” 60 dias depois. Essa “bola de neve” ia aumentando e o “emergencial” acontecia com mais freqüência do que teria ocorrido se as previsões de médio e longo prazo não tivessem sido preteridas.

Com a mudança, a quantidade de funcionários da área não aumentou; apenas houve uma realocação de parte dos integrantes das equipes locais (que eram maiores) para as equipes de longo prazo. Esta segmentação da estrutura (contando agora com uma equipe estratégica – longo prazo – e outra pra continuar cuidando do dia-dia – curto prazo – mas sem a responsabilidade de ter que prever em longo prazo) foi fundamental pra melhorar o índice de acerto das previsões.

Além disso, a divisão conseguiu melhorias operacionais significativas, já que as pessoas podem se especializar, focando a parte de análise e esquecendo a parte do dia-dia ou vice-versa, conforme o caso. Outro ponto importante é o fato de as pessoas terem perfis diferentes. Assim, o planejamento de curto prazo aloca funcionários dinâmicos, pró-ativos, aptos a “apagarem incêndios”; o planejamento de longo prazo demanda e utiliza funcionários mais analíticos, estrategistas, capazes de identificar oportunidades e de olhar a operação “de fora”.

5.2. Principais desafios operacionais

A percepção do consumidor acerca do nível de serviço da empresa ainda é muito influenciada pela memória de maus serviços da época anterior à privatização do setor de telecomunicações. A empresa – que nasceu da terceirização do SAC de seu primeiro e ainda maior cliente, uma ex-estatal do setor – carrega, por isso, um desafio ainda maior de reverter essa insatisfação, com todas as pressões e limitações inerentes ao setor (citadas em 2.1.1).

De acordo com o gerente de Planejamento de Tráfego da empresa, a principal “dor-de-cabeça” consiste em conseguir melhorar as **previsões de demanda**, tanto em termos da quantidade de ligações quanto em termos do tempo médio de atendimento. A acurácia das previsões não é boa para alguns produtos, e ele próprio “sente” que – principalmente para o produto 103 – a empresa dispõe de todas as informações que poderiam fornecer uma melhor previsão, estando apenas carente de um método mais adequado.

Outro desafio consiste em, a partir do *input* fornecido pela previsão de demanda, **dimensionar** corretamente a **capacidade de atendimento** em cada período ou, em outras

palavras, a quantidade necessária de atendentes por horário, de forma contínua durante a operação, em face de mudanças em alguns *inputs*, premissas e/ou parâmetros.

Dentro desta questão do dimensionamento, o plano de horários dos atendentes também consiste em um problema, pois eles não gostam quando seus horários, pausas para lanches, folgas e férias mudam (e muitas vezes é interessante para a Contax – do ponto de vista da eficiência operacional – fazer essas mudanças). Essa insatisfação pode causar pedidos de demissão, absenteísmo e piora na qualidade do atendimento.

Um outro ponto que revela um grande potencial de melhoria a ser estudado diz respeito ao operador multi-produto. O gerente de Planejamento de Tráfego da empresa tem a impressão que é economicamente vantajoso utilizar um mesmo operador para atender duas ou mais operações diferentes ao mesmo tempo, de modo a tentar reduzir a sua ociosidade. Isso poderia acontecer no caso de operações relativamente similares em que o mesmo operador pudesse trabalhar e que apresentassem comportamentos complementares de demanda ao longo do dia (ou da semana ou do mês).

5.2.1. Previsão de demanda

Conforme descrito na seção 5.1, a elaboração da previsão de demanda de longo prazo é realizada para cada produto da empresa por parte da respectiva coordenação. De acordo com o que foi manifestado pelo gerente de Planejamento de Tráfego (citado na seção 5.2), o produto 103 parece ser o que encerra o maior potencial de melhoria na sua previsão de demanda, a ser alcançado através de um método mais adequado para tal.

Isso ocorre porque, em relação ao produto 103, a empresa dispõe de uma boa quantidade de informação que poderia servir de subsídio para tornar a previsão um pouco melhor do que realmente costuma ser. Essas informações são basicamente as quantidades de contas enviadas para os clientes em cada data (normalmente 5-6 dias antes do vencimento) e as suas próprias datas de vencimento. O "sentimento" da gerência e da coordenação do produto 103 é que a demanda de ligações para o produto é influenciada pelos eventos de chegada da conta e do seu próprio vencimento, datas em que haveria maior acesso ao serviço por parte dos clientes. O problema é que o sistema utilizado para previsão de demanda não leva em conta essas informações, conforme será visto ainda nesta seção.

Conforme organograma apresentado na figura 16 anterior, a área que cuida da previsão de demanda do produto 103 é a Coordenação de Planejamento de Produto – Telemar & OI, que é responsável, na verdade, por toda a atividade de previsão de demanda e dimensionamento da capacidade de atendimento das operações Telemar e OI, compreendendo os produtos 102, Suporte técnico e reparo de defeitos da Telemar, Velox e OI, além do próprio 103.

Em termos de previsão de demanda, quanto mais complexa e heterogênea for a operação, e quanto mais freqüente for a ocorrência de eventos especiais, maior será a dificuldade em realizar a previsão. Os produtos Velox e OI se encaixam nesse quadro, enquanto que o 102 consiste no mais estável deles, sem dúvida, e conseqüentemente, no de maior previsibilidade.

O Velox ainda vive um momento instável, com muitas ocorrências, campanhas na televisão e uma base de clientes ainda em crescimento. A operação OI é mais dinâmica, em função da concorrência bem maior que na telefonia fixa, da ocorrência de campanhas de *marketing*, do crescimento da base de clientes (mercado em grande expansão, ainda não consolidado). Enfim, pelo fato do mercado de telefonia celular ser bem mais dinâmico do que o de telefonia fixa, sua previsão é ainda mais difícil do que a do próprio produto 103, acerca do qual a empresa tem um domínio maior.

A magnitude dos erros cometidos no processo de previsão é capaz de dar uma boa idéia da complexidade de cada operação. Até o ano de 2004, a mensuração dos erros de previsão ocorria de forma mensal; ou seja, as previsões eram feitas para cada período de meia-hora, mas, no momento de calcular o seu erro, elas eram agregadas mensalmente, e então o total era comparado com o total mensal efetivamente ocorrido no período. O problema é que poderia estar havendo um erro de previsão para mais em um determinado dia, que acabaria sendo compensando (na ótica do erro mensal) por um erro de previsão para menos em outro dia do mesmo mês, quando, na verdade, a empresa teria perdido dinheiro em um dia (devido à ociosidade) e nível de serviço no outro (devido ao sub-dimensionamento da capacidade de atendimento). Assim, o indicador do erro não era verdadeiro.

Desde o ano de 2005, então, o horizonte de mensuração do erro de previsão para esses produtos passo a ser diário (ao invés de mensal), eliminando a possibilidade de ocorrência de equívocos como os citados acima. No exemplo em questão, o método de acompanhamento mensal do erro indicaria um acerto na previsão consolidada do mês, enquanto que o

acompanhamento diário acusaria erro de previsão dos 2 dias (um para mais e outro para menos).

A tabela 4 a seguir revela a magnitude do erro absoluto percentual médio (segundo a ótica diária) cometido durante o processo de previsão de demanda para cada um dos produtos citados, em meados do ano de 2006.

Tabela 4 – Magnitude do erro absoluto (%) médio por produto, 2006

Produto/operação	Magnitude do erro médio
OI	8 a 10%
Velox	6%
103	5%
Suporte técnico e reparos	5%
102	3 a 4%

Fonte: Material coletado nas entrevistas

Na opinião dos responsáveis pela previsão de demanda, esses indicadores de erro durante o processo encontram-se em patamares baixos, sendo considerados mais do que aceitáveis. No entanto, os mesmos entendem que existe um potencial de melhoria no processo, de forma que esses erros podem se tornar ainda menores, especialmente no caso de alguns produtos.

Como pode ser observado e já havia sido colocado, a operação OI revela-se como sendo a de menor previsibilidade, devido às complexidades do mercado, enquanto que o produto 102 aparece como sendo o mais previsível, por conta, principalmente, da estabilidade da sua operação e pelo fato de não apresentar muitos fatores capazes de impactar sua demanda, que é bem mais linear.

Por causa da estabilidade da operação e por não apresentar mudanças significativas ao longo do tempo, o produto 102 foi escolhido para participar de um estudo-piloto (iniciado no começo do ano de 2006) cujo intuito é mudar a metodologia de acompanhamento do erro de previsão. A compensação de erros hipotetizada anteriormente em dois dias do mesmo mês

pode, na verdade, ser ainda mais drástica e acontecer ao longo de um dia, se um período de meia-hora tiver sua demanda superestimada enquanto outro período de meia-hora do mesmo dia for subestimado em termos de demanda.

A nova metodologia mensura o erro de meia em meia-hora (que é o bloco unitário mínimo de tempo, período para o qual são realizadas as previsões, consistindo na unidade ideal para o acompanhamento do erro). Os erros obtidos são maiores e, obviamente, mais realistas. Para o produto em questão, a meta para o erro de previsão – antes da mudança de metodologia – era de 5%. Depois da migração para a nova metodologia, a meta passou a ser de 12% de erro. Os desvios obtidos na previsão de demanda do produto 102, sob este novo horizonte de acompanhamento, têm sido considerados bem satisfatórios, portanto: na faixa de 6%, em média.

Em algumas situações atípicas, esse erro chega a subir – mas raramente ultrapassa – para um patamar de 10% (que ainda é bem aceitável, face à própria meta): Copa do Mundo (devido à ausência de histórico), meses com muitos "dias enforcados" e feriados locais), férias, problemas com grandes empresas (como no caso da Varig), entre outras.

O prazo previsto de migração dos outros produtos para este novo horizonte de acompanhamento do erro de previsão é janeiro de 2007.

A rotina para a elaboração da previsão de demanda, empreendida pela equipe da Coordenação de Planejamento de Produto – Telemar & OI, envolve as seguintes etapas:

1. Analisar e coletar dados do histórico da operação;

2. Determinar o padrão de ligações e com isto o volume de ligações para cada mês, semana, dia e para cada meia-hora do dia, usando os dados históricos da operação;

3. Projetar a carga de trabalho, que servirá de base para o cálculo da necessidade de atendentes, assim como de recursos de infra-estrutura.

Os tipos de dados a serem coletados variam de acordo com o horizonte de planejamento que está sendo feito:

- Planejamento Anual

    - Como as ligações se comportam ao longo dos meses?
    - Como o tempo de atendimento se comporta?
- Planejamento Mensal
    - Como as ligações e o TMA se comportam ao longo das semanas do mês?
    - Como as ligações e o TMA se comportam ao longo dos dias da semana?
    - Dias extraordinários (promoções, feriados)
- Planejamento Diário
    - Como as ligações e o TMA se comportam ao longo do dia?
    - Eventos (treinamentos, reuniões)
    - Ações extraordinárias (divulgação em mídia)

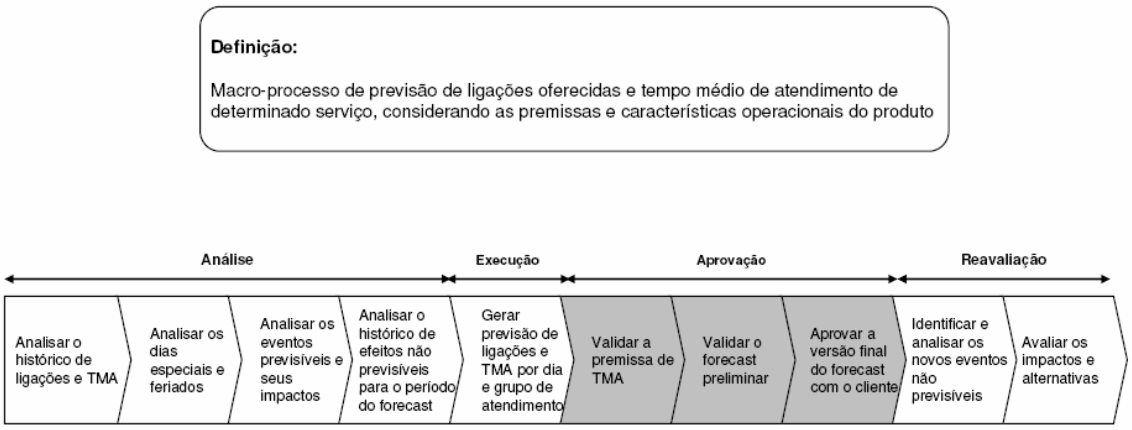
A maneira como é realizada a projeção da carga de trabalho também depende do horizonte de planejamento. Para o planejamento anual, aplica-se a tendência de crescimento (ou decréscimo) ocorrida nos últimos anos para se determinar o montante de ligações que irá ocorrer no ano em questão; os percentuais de sazonalidade são então aplicados em cada mês, para que possa ser feita a distribuição do montante anual ao longo de cada período. Outra forma consiste em aplicar a *contact rate* (taxa de ligações por cliente na base; que também sofre sazonalidade) ao total de clientes na base em cada mês.

Para o planejamento mensal, divide-se o volume mensal de ligações (projetado anteriormente) em semanas e depois em dias, utilizando os percentuais (semanais e diários) históricos para a correta alocação a cada período. Esses percentuais de sazonalidades são aplicados não somente à demanda de ligações, mas também ao tempo médio de atendimento que, da mesma forma, precisa ser previsto.

O cálculo realizado para o planejamento diário é análogo: multiplica-se o percentual histórico de ligações de cada período de meia-hora pelo volume de chamadas previsto para o dia. Um cálculo semelhante é feito para o TMA.

De uma forma geral, o processo completo de elaboração da previsão do volume de ligações e do TMA pode ser visualizado através do esquema apresentado na figura 17 a seguir.

Figura 17 – Macro-processo de previsão de ligações e TMA



Fonte: Material coletado nas entrevistas

Por causa deste processo e do diferente grau de complexidade da operação de cada produto, os prazos de antecedência necessários para a previsão da demanda dos mesmos varia[4] de acordo com a duração do processo de validação junto ao cliente e, principalmente, com o tempo necessário para selecionar, contratar e treinar os operadores (bastante variável, devido as diferentes complexidades operacionais dos produtos). Esta necessidade de contratação está sempre presente na realidade da empresa, pois a rotatividade (devido à demissão) dos atendentes situa-se na faixa de 4 a 5% ao mês; assim, mesmo se a operação for estável em termos de crescimento de base, existe a necessidade mensal de contratação para repor os demissionários. A contratação mensal só deixa de acontecer, naturalmente, nos casos em que a demanda de ligações estiver caindo.

Depois de gerada a informação de previsão, os valores são validados com os clientes internos e com o cliente externo (contratante), a necessidade de pessoas é identificada (dimensionamento) e as contratações são decididas.

Sob uma ótica menos global, o processo de previsão de demanda pode ser visto de forma mais detalhada. O primeiro passo consiste na geração de uma previsão básica, realizada pelo *software* de gerenciamento de força de trabalho, o TotalView. Um dos seus módulos é o de *forecast*, que trabalha com o histórico de demanda e faz a previsão de ligações baseado nas 13 últimas semanas, calculando uma média ponderada, com pesos diferenciados (maiores para as semanas mais recentes e menores para as mais remotas, tendo como sugestão – padrão para todos os produtos – os valores que constam na tabela 5 a seguir, mas que podem ser customizados pelo analista). A previsão é gerada para cada intervalo de meia-hora, levando em consideração também a sazonalidade do dia da semana. Assim, por exemplo, para calcular a demanda de ligações da próxima terça-feira, no período de 10:30-11:00, o TotalView considera o histórico de demanda desse mesmo intervalo de meia-hora e calcula a média ponderada da demanda das últimas 13 terças-feiras. O *software* realiza a previsão dessa maneira, não só para a quantidade de ligações, mas também para o tempo médio de atendimento (TMA).

Tabela 5 – Pesos-padrão do TotalView usados no cálculo da média ponderada de demanda

Semana	Peso
Última (1)	11%
Penúltima (2)	11%
Antepenúltima (3)	11%
4	11%
5	9%
6	9%
7	7%
8	7%
9	7%
10	5%
11	4%
12	4%
13	4%

Fonte: Material coletado nas entrevistas

O segundo passo consiste em uma crítica – por parte dos analistas da coordenação – a esses números gerados pelo TotalView. O *software* busca o histórico do dia em questão nas últimas

13 semanas, mas se nesse histórico houver alguma informação poluída (por exemplo: queda no sistema, acarretando um TMA mais baixo durante o período; ou um jogo da Copa do Mundo, acarretando redução na quantidade de ligações), a limpeza deve ser feita pelo analista. Este deve fazer uso do seu bom senso e poder analítico para, se perceber que existe algum desvio grande no histórico, realizar o ajuste manualmente – modificando ou expurgando do histórico o dado discrepante – e gerar novamente a previsão, de forma pontual. Este acompanhamento acontece diariamente.

Depois da geração da prévia da previsão (número inicial do TotalView) e da crítica inicial da equipe, o terceiro passo consiste em incorporar aos valores previstos o possível impacto causado por alguns eventos especiais (segunda-feira de carnaval ou uma ocorrência diferente em algum sábado, por exemplo). A maneira de incorporar tais impactos depende da natureza do evento (se são novos ou desconhecidos) e do prévio conhecimento das suas conseqüências. Por exemplo, a Copa do Mundo impacta diretamente a demanda no período, já que a quantidade de ligações durante os jogos do Brasil (e durante o próprio dia do jogo) apresenta queda significativa. Mas, em 2002, os jogos aconteceram de madrugada e não há histórico para a Copa de 1998; assim, para prever os impactos da Copa de 2006, foi necessário buscar eventos similares, como a Copa das Confederações e a Copa América.

Em relação ao produto Suporte técnico e reparo de defeitos, a ocorrência de chuvas caracteriza-se como um evento especial, já que impacta diretamente a quantidade de clientes com terminais telefônicos defeituosos e, conseqüentemente, a demanda de ligações para o *call center*. No entanto, este evento torna-se um pouco mais conhecido com o uso de previsões meteorológicas por período (coletadas junto ao INMET – Instituto Nacional de Meteorologia), que servem de *input* para um ajuste na demanda prevista para a central de atendimento. Para o produto em questão, é necessário, no entanto, prever a ocorrência de chuvas com uma antecedência de 60 dias, que consiste no prazo que as previsões de demanda devem ser realizadas, de forma a viabilizar a contratação e treinamento de eventuais novos operadores. Mesmo com toda essa antecedência, o índice de acerto costuma ser bom, de acordo com o coordenador do setor. A previsão meteorológica de curto prazo, ainda mais precisa e capaz de identificar dias e horários com ocorrência de chuvas, serve de *input* para as equipes locais de planejamento de curto prazo, que cuidam do ajuste fino no dia-a-dia.

Alguns dos eventos especiais podem ter seu impacto previsto de uma forma melhor, quando a equipe de analistas é subsidiada por algumas premissas recebidas de outros setores, capazes de impactar cada um dos produtos.

Estes eventos especiais incluem, entre outros:

- campanhas publicitárias na televisão – aumentando a demanda do OI e do Velox, a quantidade de questionamentos e demanda por ligações;
- previsão de crescimento da base de clientes;
- implantação de novos serviços;
- ações de manutenção e modernização da rede – que podem gerar uma quantidade grande de problemas e defeitos;
- mudanças no procedimento (roteiro de atendimento) – que podem fazer com que o operador fale mais ou menos durante o atendimento, impactando o TMA;
- implantação do *No-changes* (um sistema diferenciado, encaixando o procedimento completo do operador em uma tela apenas, sem a necessidade de navegação) na operação – que reduz o TMA;
- mudanças na URA, que podem impactar o TMA.

A relação dessas premissas que os analistas devem levar em conta no momento de incorporar possíveis impactos à previsão de demanda encontra-se no *checklist* apresentado no Anexo I.

O ideal é que essas premissas possam ser conseguidas com a antecedência necessária para que a previsão seja acurada. E isto consiste no grande desafio, na grande "dor-de-cabeça" do setor, em termos de previsão: conseguir, dentro do prazo necessário, todos os *inputs* importantes para a geração da previsão. Em suma, na opinião da equipe, o problema maior em relação à previsão de demanda não reside no processo em si, mas na obtenção dos *inputs*.

A dificuldade na obtenção dessas informações pode ser mais bem visualizada através de um exemplo. A Coordenação de Planejamento de Longo Prazo (Telemar & OI) precisa saber, em março, qual será o período de campanha publicitária da OI para o dia das mães, que tipo de cliente vai ser impactado (pré-pago ou pós-pago), qual será a grade de mídia da campanha,

em quais dias serão veiculados anúncios na televisão; tudo isto, para que possam ser previstos os impactos nos volumes de ligações em cada dia. O que ocorre é que estas informações até estão disponíveis, mas, em função da estratégia de marketing da empresa, o setor não as recebe com suficiente antecedência (para poder gerar, dentro do prazo correto, a quantidade necessária de operadores a serem contratados e treinados). Isto acontece porque a concorrência pode contra-atacar com eficiência, se as campanhas forem divulgadas com antecedência e a informação "vazar". Este problema é ainda mais crítico nos produtos que requerem uma duração maior de treinamento dos operadores, pois o setor de Recrutamento & Seleção precisa receber a informação com mais antecedência ainda.

Por tudo isso, o contato com cliente torna-se muito importante. Uma maior interação faz com que ele comece a se co-responsabilizar e passar mais premissas e informações. Daí a importância do processo de validação da previsão de demanda junto ao cliente. Já que é em função do volume previsto que a Contax será remunerada pelo serviço, o cliente precisa validar o volume de forma a gerar garantias (o pagamento acontecerá com base naquele volume mínimo e, se ocorrer uma demanda muito maior que a prevista, a Contax não tem obrigação formal de atendê-la com perfeição).

O objetivo disso tudo não consiste em apenas dividir as responsabilidades; mas também fazer o cliente entender a importância de informar essas premissas (como informações sobre campanhas na televisão ou em revistas, planos de minutos, troca de equipamentos do Velox – reduzindo os problemas de conexão e fazendo o volume de ligações despencar – promoção do dia dos pais e pacote OI Total) para ajudar na construção da previsão de demanda.

No caso específico do produto 102, a equipe de planejamento normalmente mostra os resultados do estudo de previsão e o cliente valida facilmente, sem muitas alterações. Isso ocorre porque existem poucos fatores previsíveis capazes de impactar a demanda e o cliente acaba não tendo muita informação a fornecer.

Mas, de uma forma geral, o contato com setores internos é imprescindível, devido aos *inputs* que estes devem fornecer, fundamentais para a elaboração da previsão de demanda. É o caso do setor de Faturamento da Telemar, cuja importância no processo ficará clara no detalhamento da previsão do produto 103, ainda nesta seção.

Depois da geração da análise crítica da equipe de longo prazo e da inclusão das premissas impactantes, o quarto passo do processo de previsão consiste na validação interna, junto ao planejamento local de curto prazo. Eles têm conhecimento do cotidiano e das pessoas do dia-a-dia, podem envolver o gerente da operação, têm contato com a contratante, podem informar alguma estratégia que está sendo mudada no produto ou alguma modificação na URA. Tudo isto pode agregar valor à previsão e é importante que o setor de longo prazo seja alimentado por estas informações, de forma que a previsão esteja mais madura (mais refinada do que o número fornecido pelo TotalView) no momento de fazer a validação com o cliente.

Em relação especificamente ao produto 103 (serviços e contas de telefonia fixa) – produto identificado como aparentemente o de maior potencial de melhoria na sua previsão de demanda, conforme mencionado anteriormente – a previsão é elaborada mensalmente, sempre com 45 dias de antecedência em relação à necessidade de contratação de operadores. Este prazo se justifica, para este produto, por causa do processo de validação com o cliente, do tempo de seleção do novo funcionário e dos 30 dias necessários para o treinamento do operador nas suas futuras atividades.

De acordo com o que foi descrito anteriormente nesta seção, algumas premissas – capazes de impactar cada os produtos – recebidas de outros setores da empresa, podem subsidiar a equipe de analistas no seu trabalho de previsão de demanda.

No caso do produto 103, o lançamento do plano de minutos na telefonia fixa (dentro do qual a cobrança deixa de acontecer no formato de "pulso"), por exemplo, tem potencial para impactar o volume de ligações para o 103. A equipe de previsão – de posse da informação sobre o lançamento do plano – realiza, então, uma análise para gerar um aumento ou decréscimo na volumetria, de acordo com o impacto esperado. Esse ajuste deve se basear em um evento parecido que tenha ocorrido no passado (neste ponto, atua a intuição do analista); por exemplo, se a empresa acha que o plano de minutos vai trazer acréscimo na conta dos clientes, o analista deve buscar no histórico qual foi o impacto – no volume de ligações e no TMA – de um reajuste de tarifas.

Essas premissas também envolvem o comportamento dos ciclos das contas telefônicas dos clientes, uma informação oriunda do setor de faturamento da Telemar e que ajuda a previsão de demanda do produto 103. Essa informação consiste, basicamente, nas quantidades de contas enviadas para os clientes em cada data (normalmente 5-6 dias antes do vencimento).

Ela é útil porque a demanda de ligações para o produto é influenciada – segundo a equipe de previsão – pelos eventos de chegada da conta no destino e do seu próprio vencimento, datas em que haveria maior acesso ao serviço por parte dos clientes, para fins de esclarecimento e/ou reclamação.

Na verdade, durante todo o período durante o qual a conta está com o cliente (4-5 dias desde a chegada até o vencimento, sendo a véspera e a antevéspera as datas mais críticas depois da chegada e do vencimento), ocorre um acréscimo no volume de ligações, mas o maior impacto acontece mesmo no momento da chegada da conta na residência do cliente, na opinião da coordenação responsável pelo produto. A conta chega ao destino ao longo do dia, mas é no início da própria tarde e na parte da manhã do dia seguinte, quando mais clientes estão com a conta em mãos, que a quantidade de ligações aumenta significativamente. Segundo os analistas, o impacto pós-vencimento (quando o cliente que esqueceu de pagar a conta liga para saber como proceder) existe, mas é pequeno, não sendo significativo em termos de aumento de demanda.

O problema é que o TotalView não leva em conta isso; a sazonalidade do dia da semana é considerada, mas não a distância do dia em relação à chegada da conta ou ao seu vencimento. Por exemplo, a conta de um determinado cliente vence no dia 10 de cada mês, e não na segunda quarta-feira do mesmo; então, no mês em que o dia 10 caísse em uma terça-feira, esse dia da semana deveria ser mais impactado do que os outros, enquanto que no mês em que o dia 10 caísse em uma quarta-feira, o impacto maior deveria ser incluído neste dia. Mas o TotalView considera, em termos de sazonalidade, todas as terças-feiras da mesma forma, seguindo o mesmo raciocínio para os outros dias da semana. Além disso, há meses em que as contas não são enviadas exatamente 5 dias antes do vencimento, e este outro evento especial (chegada da conta ao destino) também deve ser considerado.

Essas informações são levadas em conta no processo de previsão, mas não de forma automática. Toda a análise é feita pela equipe de previsão – e não pelo *software*! – que:

1. recebe e avalia o cronograma da área de faturamento (conforme exemplificado para o mês de maio de 2006 na tabela 6 a seguir);

2. verifica no histórico o impacto dos eventos de chegada e vencimento das contas, para o ciclo (grupamento de contas em cada filial) em questão;

3. acrescenta – por fora do TotalView – o impacto histórico desses eventos (exemplo: 5% e 3%, respectivamente) ao volume inicialmente previsto pela ferramenta para esses dias;

4. critica os resultados;

5. e retro-alimenta o *software*, ajustando a previsão que havia sido gerada.

Durante esse processo, os analistas não levam em conta um possível impacto da véspera e da antevéspera do vencimento da conta e nem do período pós-vencimento; apenas o efeito dos eventos chegada e vencimento.

A interação da equipe de previsão com o setor de faturamento, neste momento, acontece nos dois sentidos, recebendo *input* (exemplificados na tabela 6 a seguir), mas também fornecendo *output*, na forma de opinião a respeito da melhor data de postagem das contas, de acordo com o impacto operacional no *call center*, conforme será explicado a seguir.

A conta precisa ser enviada ao cliente com, no mínimo, 5 dias de antecedência em relação à data do vencimento (segundo regulamentação da ANATEL). De uma maneira geral, a chegada ocorre 1 dia após a postagem.

Tabela 6 – Cronograma de envio e vencimento de contas do faturamento, maio de 2006

Ano	Mês	Data de Postagem	Vencimento	Ciclo	Contas
2006	Mai	17/04/06	23/04/06	RJ - 5	260.710
2006	Mai	24/04/06	01/05/06	RJ - 6	634.953
2006	Mai	25/04/06	04/05/06	RJ - 7	587.118
2006	Mai	27/04/06	07/05/06	RJ - 8	589.031
2006	Mai	02/05/06	11/05/06	RJ - 9	703.452
2006	Mai	08/05/06	13/05/06	RJ - 1	469.345
2006	Mai	08/05/06	15/05/06	RJ - 2	506.432
2006	Mai	09/05/06	17/05/06	RJ - 3	198.896
2006	Mai	17/04/06	23/04/06	MG - 5	201.826
2006	Mai	24/04/06	01/05/06	MG - 6	416.176
2006	Mai	25/04/06	04/05/06	MG - 7	458.682
2006	Mai	27/04/06	07/05/06	MG - 8	515.223
2006	Mai	02/05/06	11/05/06	MG - 9	685.843
2006	Mai	08/05/06	15/05/06	MG - 10	581.568
2006	Mai	08/05/06	13/05/06	MG - 1	110.847
2006	Mai	09/05/06	17/05/06	MG - 3	310.536
2006	Mai	17/04/06	23/04/06	CE - 5	142.251
2006	Mai	24/04/06	01/05/06	CE - 6	680.389
2006	Mai	25/04/06	04/05/06	CE - 7	533.902
2006	Mai	27/04/06	07/05/06	CE - 8	626.637
2006	Mai	02/05/06	11/05/06	CE - 9	419.440
2006	Mai	08/05/06	15/05/06	CE - 10	218.537
2006	Mai	08/05/06	13/05/06	CE - 1	218.438
2006	Mai	08/05/06	15/05/06	CE - 2	150.119
2006	Mai	09/05/06	17/05/06	CE - 3	245.752
2006	Mai	17/04/06	23/04/06	BA - 5	104.650
2006	Mai	24/04/06	01/05/06	BA - 6	404.679
2006	Mai	25/04/06	04/05/06	BA - 7	562.513
2006	Mai	27/04/06	07/05/06	BA - 8	188.029
2006	Mai	02/05/06	11/05/06	BA - 9	368.817
2006	Mai	09/05/06	17/05/06	BA - 3	45.959

Fonte: Material coletado nas entrevistas

Em relação a esta data de postagem, existe um *trade-off*: quanto mais postergada for essa data, mais o cliente vai demorar a pagar e mais impactado ficará o fluxo de caixa da Telemar; por outro lado, se essa data for muito antecipada, o momento de corte na conta do cliente terá que acontecer muito cedo (até por que existe um prazo necessário para a execução das outras atividades pré-postagem, como o processamento da conta), e o faturamento do mês em questão acaba sendo prejudicado, pois incluirá menos tempo de utilização do serviço por

parte do cliente. De uma maneira geral, esse segundo efeito é mais forte e a empresa acaba preferindo deixar pra "cortar" a conta do cliente o mais próximo possível do dia de vencimento sem, naturalmente, prejudicar o prazo de antecedência regulamentado pela ANATEL (5 dias).

Mas, dentro dessa "janela" criada pelo *trade-off*, o setor de faturamento convida a equipe de previsão a participar da decisão, oferecendo duas opções de data de postagem (dois dias úteis seguidos). Do ponto de vista do faturamento, sempre uma das datas é melhor – de acordo com os aspectos descritos anteriormente no *trade-off* – mas a equipe de previsão sinaliza qual das duas minimizaria o impacto operacional no *call center* (em outras palavras, a data de postagem que faria com que a chegada das contas ao destino acontecesse no dia em que a central estivesse mais ociosa).

Atualmente, a previsão básica é extraída do TotalView e a coordenação realizada toda essa análise no Excel. A opinião geral é que as informações do faturamento poderiam ser usadas de uma forma mais sistemática para poder ser construído um método de previsão mais acurado, levando em conta todos esses aspectos e não apenas a sazonalidade do dia da semana, como faz o TotalView. A idéia é fugir mais dessa pós-análise não-metódica no Excel – que depende muito da intuição do analista – e poder contar uma ferramenta que apóie a equipe neste momento, de forma que se possa evitar o trabalho manual – que aumenta a possibilidade de erro.

A respeito do tempo médio de atendimento, o TMA, sua magnitude tem variado muito ao longo do dia, assumindo, na parte da tarde, valores 35-40 segundos maiores que os tempos matutinos, segundo a coordenação da área. A impressão geral é que isso ocorre, principalmente, pela maior concentração de operadores novos na parte da manhã que vem ocorrendo recentemente. Durante a noite, o TMA é ainda mais alto, principalmente – na opinião da equipe de previsão – por causa do menor nível de serviço oferecido neste período, o que pode fazer com que o cliente aguarde mais tempo na fila e reclame a respeito com o atendente antes de começar a tratar do assunto principal, propriamente dito, aumentando o tempo de atendimento.

Fisicamente, existem centrais de atendimento do 103 no Rio de Janeiro, em Minas Gerais e no Ceará. Mas, em termos organizacionais, tudo funciona como se houvesse uma central única de atendimento, abrangendo potencialmente os clientes de todos os estados. Na verdade, o

cliente de Belo Horizonte disca o mesmo número (103) do que o de Salvador, por exemplo, e é direcionado para a central responsável (Minas Gerais, no caso); mas, se não houver disponibilidade na central em questão, a ligação é transferida para o local onde será mais rapidamente atendida, com o menor tempo de espera. É mais interessante para a empresa (ganhos de escala) e para o cliente (atendimento mais rápido) trabalhar com o atendimento consolidado, onde o mesmo operador é capaz de atender todos os clientes da rede. Para prever a demanda, já que esta se superpõe para todas as centrais, torna-se necessário apenas elaborar uma previsão consolidada (para todo o Brasil), como se houvesse uma central única de atendimento.

5.2.2. Dimensionamento da capacidade de atendimento

O dimensionamento consiste no estudo para adequar as estruturas físicas, técnicas e de pessoal do *call center* aos objetivos da operação de atendimento ao cliente.

A partir do momento em que foi gerada a previsão diária de demanda e esse volume – em conjunto com o TMA – foi validado junto ao cliente, começa o processo de dimensionamento da capacidade de atendimento para cada produto. Somente durante tal etapa é que a previsão de volume será detalhada em um nível intradiário, com a aplicação das curvas (construídas a partir de dados históricos) de perfil de demanda ao longo do dia.

Assim como ocorre em todo o processo, a qualidade do dimensionamento da capacidade de dimensionamento vai depender da qualidade dos *inputs* deste processo: o volume de ligações e o tempo médio de atendimento previstos para cada dia.

Em função disso e na opinião da Gerência de Planejamento de Tráfego, o produto 102 é o ideal para atacar o problema do dimensionamento, visto que sua demanda é a mais previsível, de forma que seja possível mensurar isoladamente a qualidade do processo de dimensionamento (partindo da premissa que o *input* – previsão de demanda – é de boa qualidade).

Em termos de previsão de volume de ligações e do TMA, o nível de acurácia do produto 102 é muito bom comparado com outros produtos, conforme foi mencionado na seção 5.2.1. Mas existem outras características do produto que simplificam a sua operação e acabam potencializando a sua escolha para o alvo a ser atacado durante o estudo acerca do dimensionamento da capacidade de atendimento.

O período de treinamento para os atendentes do produto 102 é o mais curto de todos: uma semana. O operador precisa aprender apenas a informar o telefone procurado pelo cliente, que envolve a digitação do local e a tentativa de localização do número. Além disso, existem apenas dois tipos de clientes do produto 102: *plus* (serviço pago) e básico (serviço não pago), que são identificados automaticamente pela própria URA no momento da ligação. Quando está atendendo um cliente *plus*, o operador pode precisar realizar uma pesquisa um pouco mais aprofundada, incluindo o ramo de atividade do serviço procurado pelo cliente, por exemplo, e envolvendo até 2 ou 3 tentativas (consultas), podendo tornar o atendimento um pouco mais demorado. Mas, de uma forma geral, o treinamento é bem elementar, consistindo basicamente no aprendizado do uso do sistema e de que atitude tomar, caso a informação não seja encontrada. No final do mesmo, os funcionários avaliados como mais capazes são direcionados para atender os clientes *plus*, mas podem atender os clientes básicos, em caso de necessidade; os operadores básicos até estão aptos a atenderem os clientes *plus*, mas tal procedimento não é o ideal, só ocorrendo em casos de emergência.

O atendente do Velox, por exemplo, é um contraponto, já que tem resolver problemas, auxiliar o cliente, configurar máquinas etc.. Seu tempo de treinamento gira em torno de 20 dias.

O treinamento do atendente do 102 também é bem mais simples, por exemplo, do que o ao qual são submetidos atualmente os operadores de Suporte técnico e reparo de defeitos da Telemar. Até meados de 2005, este atendente apenas fazia a abertura de um bilhete de defeito, quando de alguma ocorrência, e mandava um técnico para a residência do cliente. Isso era muito ruim em termos de: custos (já que o processo consistia em mandar um técnico para todos os clientes que acreditavam ter um problema com a sua linha telefônica); insatisfação do cliente (por não conseguir resolver logo o seu problema); e dos indicadores da ANATEL (que registravam tais ocorrências como defeitos).

Nessa época, não havia sido estruturada a tela de suporte técnico atual, que permite o atendente fazer uma triagem junto ao cliente, ainda durante a chamada (questionando, entre outras coisas, se o telefone é sem fio, se o aparelho está ligado; acessando a central telefônica do cliente pra ver a possibilidade de fazer um reparo na central; e retornando a ligação posteriormente para checar com o cliente se o problema foi resolvido).

Essa mudança reduziu a quantidade de técnicos indo até a residência, já que passou a ser possível identificar a não ocorrência de um defeito ou até mesmo consertá-lo, com o cliente ainda na linha. Dessa forma, os custos técnicos diminuíram (menos visitas), assim como melhoraram a satisfação do cliente (resolução mais rápida do problema) e os indicadores da ANATEL (já que consertos ainda na linha não entram na estatística de defeitos). Naturalmente, aumentou o custo de atendimento e a complexidade do treinamento do operador do produto em questão; mas o *trade-off* foi vantajoso, na opinião da Gerência de Planejamento de Tráfego.

O curto prazo exigido para o treinamento dos operadores do serviço de auxílio à lista (102) reduz a antecedência com que a previsão de demanda precisa ser elaborada, tornando-a mais acurada na maioria das vezes. Além disso, esta característica do produto acaba possibilitando uma ágil reposição (dentro do próprio mês) de mão-de-obra, se ocorrer algum problema: existe um banco de operadores pré-selecionados, que – no caso de um aumento inesperado e repentino de volume – podem ser chamados, treinados em 1 semana e estar atendendo após poucos dias. Isso é impossível de acontecer em outros produtos por causa da complexidade do atendimento e do conseqüente maior tempo de treinamento.

Por outro lado, o nível do operador do 102 é inferior; recebe menor remuneração e apresenta maior rotatividade, até porque o serviço é monótono e repetitivo, gerando um desgaste maior. A qualificação dos profissionais é menor que em outros produtos, mas se determinado atendente apresentar bom perfil, pode participar do programa de rotação de cargos, de forma que não se acomode e seja capaz de manter a motivação ao vislumbrar um possível crescimento profissional na empresa (sabendo que não vai, necessariamente, atender o cliente de auxílio à lista eternamente).

No caso, como o funcionário já tem o conhecimento da operação, das funcionalidades do computador e já percebeu se aceita passar o dia inteiro atendendo ligações, fica mais fácil e rápida a sua assimilação em relação a outros produtos, causando menos problemas. Essa migração interna pode acontecer para qualquer outro produto da empresa, oferecido nas mesmas centrais que o 102: Bahia e Pernambuco.

Na Contax, o período mensal é considerado como indo do dia 16 do mês anterior até o dia 15 do próprio mês, sendo que a primeira metade corresponde ao período de baixa, com volume fraco. Assim, de uma maneira geral, a meta para o setor de planejamento consiste em

implantar o dimensionamento antes do momento de alta (segunda metade) – ou seja, até o dia 1 de cada mês – de forma que possa lidar com eventuais aumentos da demanda.

Mas o curto período de treinamento dos operadores do produto 102 permite uma agilidade também na implantação deste dimensionamento, que costuma ocorrer tranqüilamente antes do período começar (dia 16), superando a meta geral da empresa de conseguir essa implantação no meio do período (dia 1).

Em função de tudo isso, o produto 102 foi identificado pela Gerência de Planejamento de Tráfego como o de maior potencial para a abordagem do problema de dimensionamento da capacidade de atendimento. Assim, e a partir de agora, o restante desta seção vai estar focando na descrição do processo de dimensionamento do produto 102, explicando como a equipe de planejamento consegue transformar a previsão de demanda na quantidade necessária de funcionários em cada dia e horário e que ferramentas são usadas neste processo.

Antes de iniciar a descrição do processo de dimensionamento do produto 102, no entanto, é importante apontar os termos contratuais que fazem parte do *Service Level Agreement* (SLA ou acordo de nível de serviço) entre a fornecedora (Contax) e a cliente (Telemar).

O nível de serviço para este produto diz respeito ao tempo de espera do cliente final na linha telefônica, desde o momento de entrada da ligação até ela ser atendida. Em outras palavras, é o tempo que o cliente passa "na fila", ouvindo a música de espera e aguardando o operador.

Mais precisamente, no caso do 102, o nível de serviço consiste no percentual de chamadas atendidas – dentre as completadas, apenas – em até 10 segundos (ou melhor, que aguardam no máximo 10 segundos para serem atendidas). Para efeitos comparativos, outros produtos da Contax lidam com outros valores: no caso do atendimento de cartões de crédito, por exemplo, o nível de serviço consiste no percentual de chamadas atendidas em até 20 segundos.

Como apenas as ligações efetivamente atendidas entram no cálculo do nível de serviço, os abandonos não são contabilizados (nem punidos, portanto), para efeitos de nível de serviço. Mas eles são mensurados através de outro indicador (taxa de abandono) e a Contax é multada quando essa taxa ultrapassa os 2%, em algum mês. Assim, o fornecedor estaria sendo punido duas vezes se as ligações abandonadas entrassem no total das chamadas a serem atendidas em até 10 segundos.

Em vista da possibilidade de multa, evitar o abandono costuma ser visto como prioridade, um pouco em detrimento do nível de serviço, desde que esse se mantenha acima de patamares mínimos. O nível de serviço não envolve exigências formais no contrato (como a taxa de abandono), mas influencia a relação e a dignidade comerciais; ou seja, é interessante não priorizar apenas o abandono, deixando de se preocupar em manter o nível de serviço em valores decentes.

Iniciando a descrição da rotina de dimensionamento, este processo consiste de duas grandes etapas, no caso do produto 102:

- uma análise macro, onde os volumes diários de previsão de demanda são analisados e o cálculo da base de atendentes ("a quantidade certa de pessoas no lugar certo") é realizado;

- uma análise intradiária, onde o nível de detalhe aumenta para períodos de meia-hora, o número de atendentes necessários para cada um desses períodos é calculado (levando em conta detalhes como: pausas, faltas, férias, folgas) e as escalas de horários são organizadas.

Para cada uma das etapas, uma planilha do Excel funciona como mola-mestra para o processo, utilizando como metodologia de cálculo as fórmulas Erlang de teoria de filas. O processo é razoavelmente manual, visto que não ocorre via sistema.

A primeira análise consiste, basicamente, no cálculo de necessidades, conforme evidenciado na tabela 7 a seguir. Quatro planilhas precisam ser desenvolvidas – duas para cada produto (*plus* e básico), em cada uma das centrais de atendimento (Bahia e Pernambuco) – e depois consolidadas em uma única tabela de necessidades.

O cálculo da necessidade de operadores em cada dia é operacionalizado multiplicando a demanda de ligações pelo TMA do dia e dividindo este total (demanda de minutos) pelo produto do total de minutos diários pelo qual o operador é contratado (340, já que o funcionário de 6 horas é tomado como padrão) por sua taxa de ocupação (período estimado durante o qual o atendente vai estar efetivamente falando). Essa taxa de ocupação é estimada de forma diferenciada a cada dia da semana, conforme pode ser observado na tabela 8 a seguir.

Tabela 7 – Cálculo de necessidade de recursos diários para o produto 102, central BA, cliente *plus*, outubro-novembro de 2006

Dia	Volume de ligações	TMA (seg)	Necessidade de recursos
	8.456.710	32,24	
16-seg	361.538	32,00	610
17-ter	348.243	32,00	594
18-qua	339.242	32,00	585
19-qui	322.718	32,00	562
20-sex	307.903	32,00	537
21-sáb	188.494	30,00	326
22-dom	107.984	29,00	181
23-seg	359.605	32,00	607
24-ter	346.477	32,00	591
25-qua	286.022	32,00	493
26-qui	272.157	32,00	474
27-sex	259.721	32,00	453
28-sáb	159.030	31,00	284
29-dom	91.121	30,00	158
30-seg	303.496	32,00	512
31-ter	292.453	33,00	514
01-qua	301.535	33,00	536
02-qui	286.937	33,00	516
03-sex	273.836	33,00	492
04-sáb	167.674	31,00	300
05-dom	96.071	30,00	166
06-seg	319.967	33,00	557
07-ter	308.298	33,00	542
08-qua	346.008	33,00	615
09-qui	329.211	33,00	592
10-sex	314.127	33,00	565
11-sáb	192.308	31,00	344
12-dom	110.162	30,00	191
13-seg	366.810	33,00	638
14-ter	353.344	33,00	621
15-qua	344.218	33,00	612

Fonte: Material coletado nas entrevistas

Tabela 8 – Premissas para a taxa de ocupação do atendente do produto 102, por dia da semana

Dia da semana	Taxa de ocupação estimada
Segunda-feira	93%
Terça-feira	92%
Quarta-feira	91%
Quinta-feira	90%
Sexta-feira	90%
Fim-de-semana	85%

Fonte: Material coletado nas entrevistas

A tendência de queda dessa taxa ao longo da semana pode ser explicada pelo fato dos primeiros dias da semana normalmente serem os de maior demanda (ou DMM – dia de maior movimento), quando os operadores passam mais tempo ocupados, acarretando, inclusive, a necessidade de horas-extras em algumas ocasiões.

Em seguida, essa necessidade (em termos de operadores de 6 horas) é comparada com a disponibilidade atual de recursos de atendimento. Mas existem 3 tipos de operadores em termos de jornada de trabalho: os que trabalham 4 horas por dia, de segunda a sexta e em mais um dia do fim-de-semana (sábado ou domingo); os que trabalham 6 horas por dia, de segunda a sexta; e os que trabalham 6 horas por dia, de segunda a sexta e em mais um dia do fim-de-semana (sábado ou domingo). Por isso, é necessário colocar tudo em uma só base e essa padronização (ilustrada na tabela 9 a seguir) é feita utilizando o conceito de FTE, que consiste na unidade de homens-hora equivalente a 1 operador de 6 horas (assim, por exemplo, 3 operadores de 4 horas correspondem a 2 FTEs).

Durante essa comparação, no entanto, é preciso levar em conta que alguns FTEs "disponíveis" estarão em férias e que haverá uma perda natural – implementada na fórmula da coluna "Diferença" da tabela 9 a seguir – por absenteísmo (falta, licença médica ou qualquer motivo que faça com que o operador não esteja apto a atender uma chamada – como, por exemplo, um treinamento ou uma conversa com o supervisor objetivizando um *feedback*). O absenteísmo estimado é uma premissa passada pela própria operação – que precisa controlar e efetivamente controla essas ocorrências – para o setor de planejamento. A acurácia dessa premissa é muito importante e por isso a operação precisa assumir esse compromisso e estar sempre alinhando essa estimativa em conjunto com o planejamento; caso contrário, o dimensionamento pode ficar muito prejudicado.

O absenteísmo orçado pela gerência é de 3%, mas, normalmente o valor 4% é utilizado como premissa de tal grandeza, para os dias úteis e sábados. Nos domingos e feriados, essa estimativa sobe para a faixa dos 7%. Em alguns feriados regionais (como São João no Nordeste, por exemplo), uma estimativa diferenciada é utilizada: de 15 a 20%! O período das eleições também normalmente representa um problema, já que muitos funcionários são convocados para trabalhar como mesários, não comunicam a supervisão da Contax e acabam faltando. Faltas não planejadas consistem em absenteísmo; se a falta fosse comunicada com

antecedência, o planejamento não escalaria os funcionários ausentes e o absenteísmo seria bem menor.

Tabela 9 – Cálculo de necessidade de recursos diários para o produto 102, central BA, cliente *plus*, outubro-novembro de 2006

Dia	Necessidade de recursos	4h/6d	6h/5d	6h/6d	FTE	Férias	FTE - Férias	Diferença
16-seg	610	52	106	498	639	42	597	-37
17-ter	594	52	106	498	639	40	599	-19
18-qua	585	52	106	498	639	40	599	-10
19-qui	562	52	106	498	639	40	599	12
20-sex	537	52	106	498	639	38	601	40
21-sáb	326	31		299	320	27	293	-45
22-dom	181	21		199	213	18	195	1
23-seg	607	52	106	498	639	35	604	-27
24-ter	591	52	106	498	639	35	604	-11
25-qua	493	52	106	498	639	35	604	86
26-qui	474	52	106	498	639	35	604	105
27-sex	453	52	106	498	639	35	604	127
28-sáb	284	31		299	320	25	295	-1
29-dom	158	21		199	213	17	196	25
30-seg	512	52	106	498	639	39	600	64
31-ter	514	52	106	498	639	39	600	61
01-qua	536	52	106	498	639	39	600	40
02-qui	516	52	106	498	639	41	598	40
03-sex	492	52	106	498	639	41	598	82
04-sáb	300	31		299	320	26	294	-18
05-dom	166	21		199	213	18	195	15
06-seg	557	52	106	498	639	41	598	17
07-ter	542	52	106	498	639	41	598	32
08-qua	615	52	106	498	639	45	594	-45
09-qui	592	52	106	498	639	45	594	-22
10-sex	565	52	106	498	639	45	594	5
11-sáb	344	31		299	320	26	294	-62
12-dom	191	21		199	213	19	194	-10
13-seg	638	52	106	498	639	45	594	-68
14-ter	621	52	106	498	639	47	592	-53
15-qua	612	52	106	498	639	47	592	-62

Fonte: Material coletado nas entrevistas

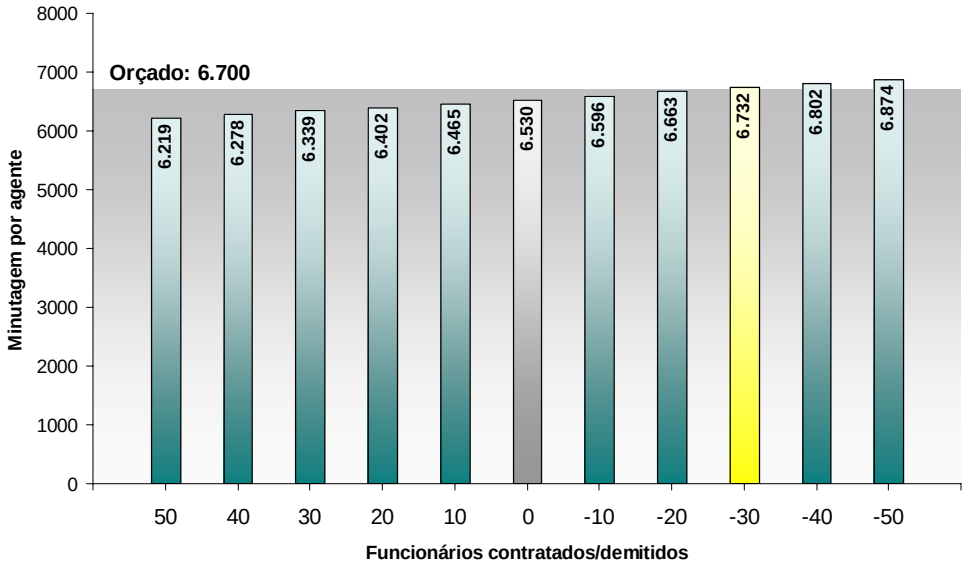
O resultado da comparação da necessidade de recursos com a disponibilidade de capacidade é um saldo ou um déficit de mão-de-obra para cada dia do mês planejado (coluna "Diferença" da tabela 9 anterior). Se houver déficit em alguns dias e sobra em outros, o planejamento pode fazer uso da escala móvel, que consiste em dar folga para funcionários nos dias com sobra (normalmente sextas-feiras) e pedir para eles cumprirem 2 horas adicionais nas segundas, terças e quartas-feiras (dias de maior movimento), totalizando as 6 horas folgadas nas sextas-

feiras. Existe uma adesão de até normalmente 80% para a escala móvel; ou seja, até 80% dos atendentes aceitam folgar na sexta-feira e trabalhar 2 horas a mais de segunda a quarta-feira).

Se a escala móvel neutralizar o eventual déficit, pode ser até necessário demitir alguns funcionários, por causa do excesso de capacidade. Isso não necessariamente precisa se concretizar, pois já existe uma rotatividade mensal natural da equipe. Dependendo do tamanho do excesso de capacidade, pode ser necessário apenas não repor as pessoas que pediram demissão. Caso a escala móvel não seja capaz de suprir o déficit, a contratação é a única saída. Toda essa análise é feita de forma consolidada, já que não há a necessidade de um dimensionamento para cada central, uma vez que os operadores da Bahia podem atender os clientes que seriam direcionados para a central de Pernambuco e vice-versa.

Nesse momento de decisão a respeito de contratação ou demissão, a produtividade dos funcionários (minutagem por agente) precisa também ser levada em conta. Com o volume de ligações e o TMA previstos para o mês, é possível determinar a minutagem por agente em diferentes cenários (com diferentes números de atendentes contratados ou demitidos), conforme ilustrado na figura 18 a seguir.

Figura 18 – Minutagem por agente para cenários com diferentes efetivos de atendentes, produto 102, consolidado Brasil, abril de 2006



Fonte: Material coletado nas entrevistas

A linha cinza representa a meta orçada pela gerência para esse indicador desempenho e o analista de planejamento deve procurar superá-la, migrando – no exemplo ilustrado na figura 18 anterior – do contingente atual (barra cinza) para um cenário com 30 funcionários a menos (barra amarela).

Finalmente essa primeira macro-análise termina com o envio – por parte do planejamento e com destino à operação – de um resumo da tomada de decisão envolvendo a forma de trabalho para o mês subseqüente, juntamente com as premissas consideradas em relação aos atendentes e feriados (pois, às vezes, o absenteísmo ou a taxa de ocupação podem estar, por exemplo, sendo subestimados para aquele mês ou um feriado local pode estar sendo esquecido – e tais informações podem ser corrigidas pelo pessoal da operação).

Esse resumo do planejamento de trabalho para o mês seguinte nada mais é do que a planilha apresentada na tabela 9 anterior, só que desta vez preenchida com as quantidades sugeridas de funcionários para o período subseqüente, ao invés do efetivo atual de atendentes. Ele é apresentado para a operação, mostrando o cenário por central (Bahia e Pernambuco), por produto (*plus* e básico) e também o consolidado, antes do ajuste da escala móvel. Outro resumo similar é apresentado em seguida, considerando a possibilidade de usar a escala móvel e assinalando os seus impactos.

O *output* da primeira análise macro é a quantidade de operadores que precisam ser contratados ou demitidos para os números desejados serem atingidos.

A partir do momento em que a decisão de contratação ou demissão foi tomada e implementada, a equipe de planejamento pode partir para a análise micro – o dimensionamento intradiário.

Nesse momento, é importante ficar claro que o dimensionamento deve ser feito para um dia apenas e esse formato-padrão repetido para os demais dias do mês. Essa exigência existe porque o horário de cada funcionário deve ser o mesmo em todos os dias do mês; ele não pode ser escalado para trabalhar de manhã em um dia e de tarde em outro dia, por exemplo. Antes de o mês começar (mas não no seu decorrer), algumas mudanças podem ser feitas nos horários dos funcionários, mas elas ficarão valendo para o mês inteiro e, de preferência, por um período até mais longo. É possível, entretanto, conseguir um pouco mais de flexibilidade em relação a isso através do uso da escala móvel, conforme explicado anteriormente.

Em vista disso, um volume de ligações e um TMA (grandezas necessárias para o dimensionamento) específicos precisam ser escolhidos para serem usados como padrão para o dimensionamento de todos os dias do mês. Normalmente, se a diferença entre o quinto DMM (quinto dia de maior movimento, dentro do mês) e o primeiro DMM não for muito grande (poucas dezenas de milhares de chamadas), o dia escolhido para o padrão é o quinto DMM, juntamente com o seu respectivo TMA.

Nessa situação, o dimensionamento vai garantir o nível de serviço desejado para esse dia e todos os dias menos movimentados, mas não para os 4 dias de maior demanda, quando haverá perda de nível de serviço. Isso não representa um problema porque o SLA que consta no contrato do produto 102 exige um específico nível de serviço mensal, e não diário. Se a exigência fosse a respeito do nível de serviço diário (como em outros produtos), o dimensionamento teria que ser feito com base no primeiro DMM, resultando em ociosidade em boa parte do mês.

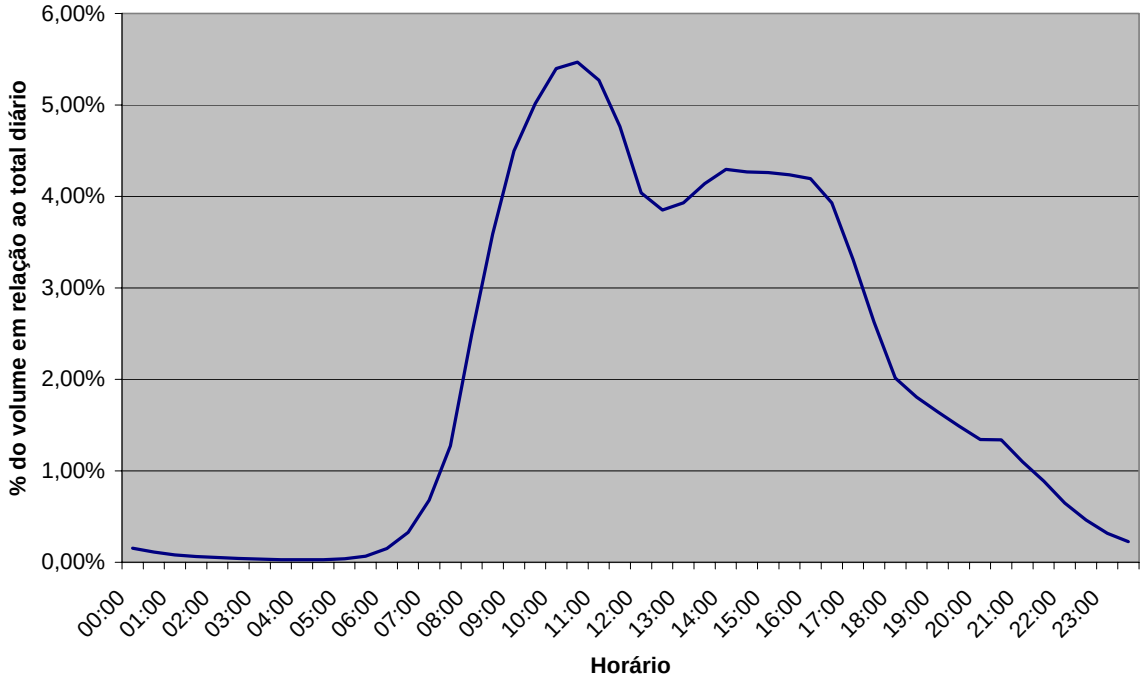
Mas se a variação entre o primeiro e o quinto DMMs for muito grande, é preciso ter cuidado, já que tal dimensionamento pode acarretar um nível de serviço muito baixo nos dias mais movimentados. Em tais situações, normalmente o padrão considerado para o dimensionamento consiste em uma média entre os 5 primeiros DMMs para o volume de chamadas e em uma média dos TMAs dos 5 primeiros DMMs, para o tempo médio de atendimento.

Em cima do dia escolhido como padrão de dimensionamento para o mês, é aplicada uma curva que reflete o perfil intradiário da demanda, ou seja, que percentual do volume diário acontecerá na primeira meia-hora do dia, na segunda meia-hora, ..., e na última meia-hora do dia. Essa curva é obtida com base no histórico de ligações recebidas a cada intervalo de meia-hora, para cada dia da semana. No caso do produto 102, as curvas de cada dia da semana são muito parecidas (principalmente de segunda à quarta-feira, com pequeno aumento vespertino no volume da quinta e da sexta-feira), sendo as do sábado e do domingo um pouco diferentes. A figura 19 a seguir ilustra a curva histórica da terça-feira.

O dimensionamento de alguns produtos com demanda mais instável considera apenas uma curva para os dias de semana, outra para sábado e outra para domingo. Devido à maior previsibilidade do 102, é possível descer ao nível de detalhe de trabalhar com uma curva diferente para cada dia da semana. Assim, se o dia escolhido como padrão de

dimensionamento for uma terça-feira, por exemplo, aplica-se a curva de terça-feira ao volume de ligações do dia em questão.

Figura 19 – Perfil histórico do comportamento intra-diário da demanda do produto 102, terça-feira



Fonte: Material coletado nas entrevistas

O resultado deste processo é uma previsão de demanda de ligações (volume e TMA) para cada meia-hora. Com o auxílio de fórmulas Erlang do suplemento de Excel chamado "Turbotab" (conforme será explicado ainda nesta seção), a planilha apresentada na tabela 10 a seguir calcula a quantidade necessária de operadores (desconsiderando e considerando o absenteísmo) para atender a demanda de cada período com um nível mínimo de serviço pré-estabelecido (normalmente 85% das ligações sendo atendidas em até 10 segundos).

Tabela 10 – Cálculo da necessidade de operadores para cada período de 30 minutos, produto 102 (*plus*), consolidado Brasil, agosto de 2006

Horário	Volume	T.M.A.	Erlangs	Oper. Nec. S/ ABS	Oper. Nec. C/ abs
00:00	586	29	9,54	13	14
00:30	430	29	7,00	10	10
01:00	308	29	5,01	8	8
01:30	243	29	3,96	6	6
02:00	196	28	3,07	5	5
02:30	156	28	2,46	5	5
03:00	127	27	1,92	4	4
03:30	106	30	1,78	4	4
04:00	108	28	1,69	4	4
04:30	106	29	1,72	4	4
05:00	145	29	2,36	5	5
05:30	253	31	4,40	7	7
06:00	576	29	9,31	12	13
06:30	1.245	31	21,32	25	26
07:00	2.572	32	45,79	50	52
07:30	4.833	32	86,00	91	95
08:00	9.446	32	168,04	173	180
08:30	13.610	32	241,96	247	257
09:00	17.059	32	303,29	310	323
09:30	19.016	32	341,52	349	364
10:00	20.483	32	364,06	371	386
10:30	20.738	32	368,83	376	392
11:00	19.999	32	351,77	357	372
11:30	18.067	32	318,14	324	338
12:00	15.320	32	272,52	278	290
12:30	14.619	32	260,15	265	276
13:00	14.904	33	270,54	276	288
13:30	15.704	33	290,84	296	308
14:00	16.293	33	296,05	303	316
14:30	16.185	33	299,75	307	320
15:00	16.157	33	293,85	301	314
15:30	16.073	33	297,68	305	318
16:00	15.903	33	294,53	300	313
16:30	14.907	33	270,98	276	288
17:00	12.551	32	222,69	228	238
17:30	9.961	32	175,02	180	188
18:00	7.647	31	133,04	138	144
18:30	6.859	31	118,30	123	128
19:00	6.243	32	110,27	115	120
19:30	5.645	31	96,52	101	105
20:00	5.095	32	89,82	96	100
20:30	5.073	32	90,92	97	101
21:00	4.171	32	74,99	81	84
21:30	3.371	32	60,71	66	69
22:00	2.456	33	45,49	50	52
22:30	1.751	31	30,47	34	35
23:00	1.200	31	20,89	25	26
23:30	851	27	12,89	16	17

Fonte: Material coletado nas entrevistas

Essas informações são, então, cruzadas com o contingente atual (utilizado no mês anterior) de atendentes, que é apresentado em uma planilha enviada pela operação (e ilustrada na tabela 11 a seguir) contendo a quantidade de operadores que estão iniciando a sua jornada a cada período do dia (resumo das escalas), tanto para os funcionários que trabalham 4 horas por dia, como para aqueles que trabalham 6. Em função do horário de início de cada agente, a planilha calcula quantos atendentes estarão disponíveis em cada intervalo de meia-hora.

Tabela 11 – Necessidade (total) e disponibilidade (4 horas) de agentes por período, produto 102 (*plus*), consolidado Brasil, agosto de 2006

Horário	Nec.c/ abs	4 Horas	06:30	07:00	07:30	08:00	08:30	09:00	09:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00
00:00	14	0														
00:30	10	0														
01:00	8	0														
01:30	6	0														
02:00	5	0														
02:30	5	0														
03:00	4	0														
03:30	4	0														
04:00	4	0														
04:30	4	0														
05:00	5	0														
05:30	7	0														
06:00	13	0														
06:30	26	0														
07:00	52	7	0	7												
07:30	95	39	0	7	32											
08:00	180	47	0	7	32	8										
08:30	257	86	0	7	32	8	39									
09:00	323	102	0	7	32	8	39	16								
09:30	364	102	0	7	32	8	39	16								
10:00	386	102	0	7	32	8	39	16	0							
10:30	392	102		7	32	8	39	16	0	0						
11:00	372	95			32	8	39	16	0	0	0					
11:30	338	63				8	39	16	0	0	0	0				
12:00	290	55					39	16	0	0	0	0	0			
12:30	276	16						16	0	0	0	0	0	0		
13:00	288	12							0	0	0	0	0	0	0	12
13:30	308	43								0	0	0	0	0	0	12
14:00	316	93									0	0	0	0	0	12
14:30	320	103										0	0	0	0	12
15:00	314	103											0	0	0	12
15:30	318	103												0	0	12
16:00	313	103													0	12
16:30	288	103														12

Fonte: Material coletado nas entrevistas

Depois de consolidar as disponibilidades dos agentes de 4 horas e dos agentes de 6 horas, um gráfico – como o exemplificado na figura 20 a seguir – sumariza a relação entre as quantidades disponível e necessária de operadores para cada período de 30 minutos.

Em cima do atual escalonamento é que a equipe de planejamento vai trabalhar para modificar a disponibilidade de operadores em cada período do dia, de modo a atingir o nível de serviço desejado.

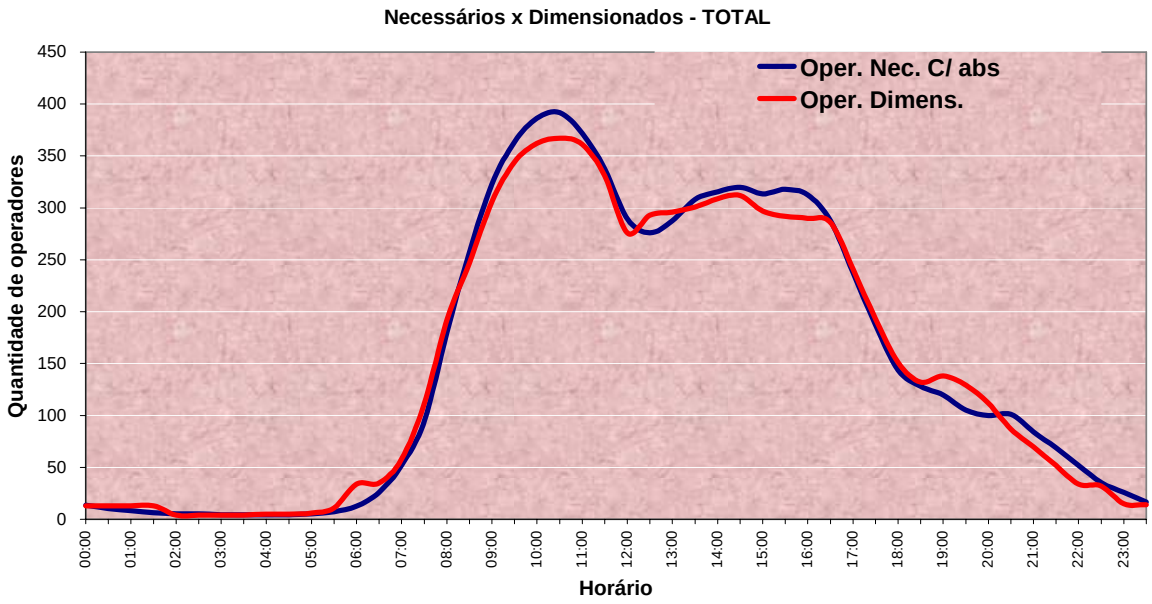
Assim, esta etapa de organização de escalas tem a função de garantir que haja a quantidade certa de pessoas em cada horário. Ela é feita através do processo de tentativa-e-erro, durante o qual é preciso analisar diversos fatores, tais como: carga horária diária, aspectos da legislação trabalhista, acordos sindicais, espaço físico disponível (PAs). Normalmente, o desejado é que a escala esteja adequada à demanda para atingir uma determinada meta. De uma forma geral, a escala pode ser feita de duas formas:

- **Cobertura Total –** onde não é permitido que haja menos agentes do que o necessário em momento algum, garantindo que a curva de demanda seja completamente coberta.

Este tipo de escala tende a aumentar a quantidade de atendentes contratados e a gerar ociosidade em determinados períodos, gerando, no entanto, melhores resultados no tocante ao nível de serviço;

- **Balanceada –** onde é assumido que, em determinados períodos do dia, a quantidade de atendentes escalados será menor que a necessária (*understaffing*), enquanto que em outros esta quantidade será maior (*overstaffing*), garantindo assim que a meta seja atingida, não de forma pontual, mas ao longo do dia.

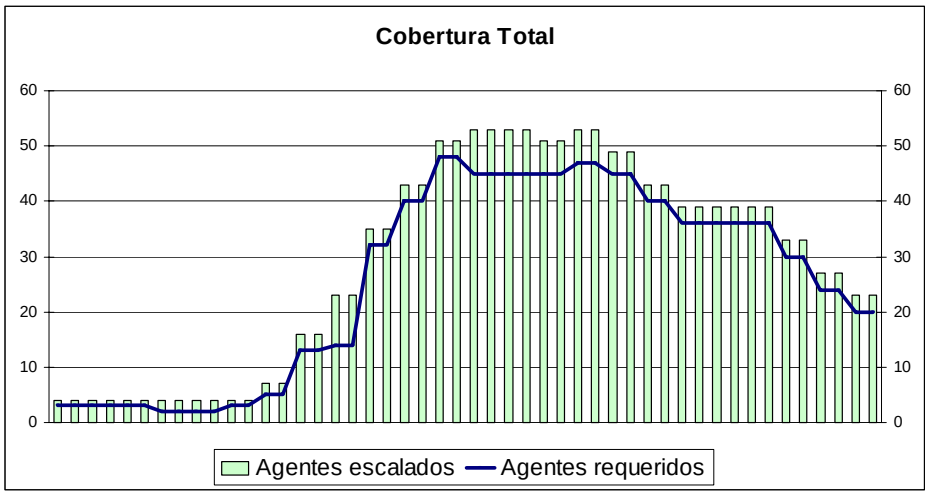
Figura 20 – Necessidade e disponibilidade consolidada (4 e 6 horas) de agentes por período, produto 102 (*plus*), Brasil, agosto de 2006



Fonte: Material coletado nas entrevistas

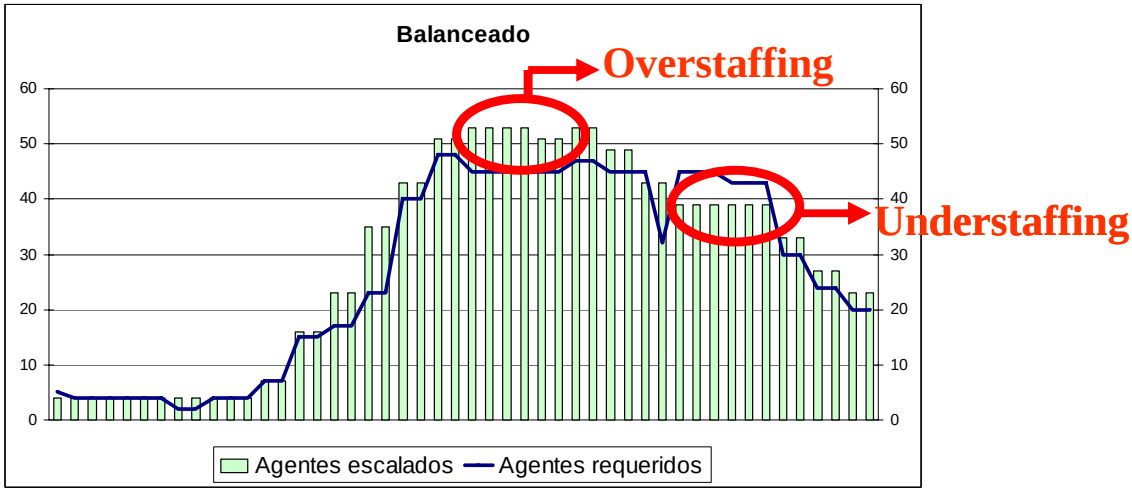
As figuras 21 e 22 a seguir ilustram graficamente o comportamento do escalonamento com cobertura total e o do escalonamento balanceado, respectivamente, em comparação com a demanda de recursos.

Figura 21 – Agentes escalonados e requeridos ao longo do dia, exemplo de escalonamento com cobertura total



Fonte: Material coletado nas entrevistas

Figura 22 – Agentes escalonados e requeridos ao longo do dia, exemplo de escalonamento balanceado



Fonte: Material coletado nas entrevistas

Para o produto 102, o escalonamento balanceado pode ser usado, já que o importante para efeitos comerciais é o nível de serviço médio do dia, e não detalhado de forma intradiária.

Durante o processo de escalonamento, a equipe de planejamento faz tentativas – com o objetivo de buscar o melhor cenário de distribuição dos recursos por horário capaz de atender o volume de chamadas previsto e validado anteriormente – modificando a quantidade de

operadores que iniciam a sua jornada a cada horário (células amarelas dispostas diagonalmente na tabela 11 anterior) e também os horários de lanche (para funcionários com jornada diária de 6 horas, respeitando os limites legais que exigem que nenhum funcionário seja escalado para lanchar menos de 1,5 hora depois do início da jornada e nem menos de 1 hora antes do fim da mesma).

É importante ressaltar que não é possível, neste momento, alterar a quantidade de operadores que vão trabalhar no mês e, conseqüentemente, em cada dia (grandeza validada anteriormente); existe liberdade apenas para alterações na alocação dos mesmos durante o dia.

Essas mudanças alteram, conseqüentemente, a quantidade disponível de atendentes em cada intervalo de meia-hora e permitem o analista lidar com flexibilidade na questão de ociosidade *versus* sub-dimensionamento. A planilha cruza, então, essa informação – de forma consolidada (operadores de 4 e 6 horas) – com a demanda de ligações anteriormente prevista e com a quantidade necessária de atendentes previamente calculada para cada um desses intervalos e estima o nível de serviço para cada período de meia-hora e o conseqüente nível de serviço médio estimado para o dia, conforme ilustrado na tabela 12 a seguir.

O cálculo deste nível de serviço estimado é realizado por intermédio de um suplemento do Excel chamado "Turbotab". Quando instalado, ele implementa funções que são usadas na planilha para efetuar o cálculo em questão, a partir da inclusão das premissas (previsão de volume e de TMA, absenteísmo e critérios para cálculo do nível de serviço – no caso, chamadas atendidas em até 10 segundos). O suplemento consiste em uma planilha de dimensionamento baseada nos princípios de Erlang.

Como pode ser observado, a teoria de filas é utilizada para o dimensionamento de agentes. Os clientes que chegam e não encontram atendentes disponíveis, aguardam <u>indefinidamente</u> em uma fila até serem atendidos (não existe abandono ou sinal de ocupado), segundo a teoria. Os tempos de espera são previstos partindo da premissa que o intervalo entre chegadas e o tempo de atendimento seguem distribuições exponenciais, e levando em consideração três fatores: o número de agentes; o número de clientes em espera na fila; e o tempo médio de atendimento.

Tabela 12 – Necessidade e disponibilidade de agentes e nível de serviço estimado por período, produto 102 (*plus*), Brasil, agosto de 2006

Horário	Nec.c/ abs	4 Horas	6 Horas	Total	Saldo de agentes	Nível de serviço
08:00	180	47	139	186	6	98,8%
08:30	257	86	175	261	4	96,8%
09:00	323	102	218	320	(3)	77,9%
09:30	364	102	250	352	(12)	0,0%
10:00	386	102	268	370	(16)	0,0%
10:30	392	102	278	380	(12)	0,0%
11:00	372	95	283	378	6	98,7%
11:30	338	63	286	349	12	99,9%
12:00	290	55	246	301	11	99,9%
12:30	276	16	288	304	28	100,0%
13:00	288	12	291	303	16	100,0%
13:30	308	43	274	317	9	99,4%
14:00	316	93	232	325	9	99,8%
14:30	320	103	218	321	1	95,8%
15:00	314	103	204	307	(7)	28,1%
15:30	318	103	195	298	(20)	0,0%
16:00	313	103	193	296	(17)	0,0%
16:30	288	103	184	287	(1)	82,5%
17:00	238	91	149	240	3	95,5%
17:30	188	60	133	193	6	98,7%
18:00	144	10	147	157	13	100,0%
18:30	128	5	138	143	15	100,0%
19:00	120	8	125	133	13	100,0%
19:30	105	12	111	123	18	100,0%
20:00	100	25	91	116	16	100,0%
20:30	101	25	66	91	(10)	0,0%
21:00	84	25	50	75	(9)	0,0%
21:30	69	25	36	61	(8)	0,0%
22:00	52	25	18	43	(9)	0,0%
22:30	35	20	16	36	1	89,2%
23:00	26	17	12	29	3	98,6%
23:30	17	13	11	24	7	100,0%
PA´s	392			380		66,19%

Fonte: Material coletado nas entrevistas

É bom deixar claro que, até esse momento, ainda não estão sendo levadas em conta as possibilidades de uso de recursos capazes de melhor o nível de serviço em cada horário, como horas-extras e o banco de horas proveniente da escala móvel. Ou seja, os níveis de serviço ainda podem aumentar e os patamares desejados não precisam ser necessariamente atingidos já neste estágio do processo de dimensionamento, que ainda não reflete o uso de tais recursos.

Tendo a consciência, então, deste provável aporte de mão-de-obra adicional, é interessante distribuir as lacunas de recursos em três momentos do dia (manhã, tarde e noite), para que depois possam ser preenchidas uniformemente pelos recursos adicionais. Talvez fosse

intuitivo deixar um déficit maior para a parte da noite, na esperança de haver uma maior adesão de horas-extras e compensação de banco de horas nesse período, mas não é isso que acontece na Contax, conforme será detalhado ainda nesta seção.

Voltando à etapa de organização das escalas, vale dizer que consiste em um processo iterativo, onde a principal motivação do analista – no momento de decidir quantos atendentes devem iniciar sua jornada em cada horário – é maximizar o nível de serviço médio do dia. Esse objetivo em relação ao nível de serviço pode ser mais ou menos agressivo dependendo de alguns fatores, como, por exemplo, o produto em si (básico – menos agressivo – ou *plus* – mais agressivo) e se no dia em questão está programada uma medição da ANATEL (mais agressivo).

O nível de serviço em cada faixa de horário, em si, não costuma ser uma preocupação para o analista, já que o importante é elevar o nível de serviço médio do dia. Em cada faixa de horário, entretanto, existe uma preocupação para não deixar um déficit muito grande de atendentes escalados, em relação à capacidade demandada.

Conforme foi dito anteriormente, o ideal é distribuir uniformemente esse déficit ao longo de três períodos do dia (manhã, tarde e noite) a fim de que possa ser reduzido (e até eliminado) pelo aporte dos recursos provenientes das horas-extras e do banco de horas e de que não haja impacto muito forte em nenhum momento.

A preocupação com os déficits intradiários existe porque, nos horários com muita carência de atendentes, pode haver grande incidência de abandonos. E isso pode ser muito ruim por dois motivos. Em primeiro lugar, porque aumenta a chance de a Contax ter que pagar uma multa por excesso de ligações abandonadas.

E em segundo lugar, porque o cliente que abandona uma chamada normalmente liga novamente um pouco mais tarde; se demorar a ser atendido novamente, ele pode abandonar outra ligação, e mais uma ou duas. Na maioria das vezes, no entanto, ele vai insistir até ser atendido já que, no caso do 102, ele provavelmente está precisando daquela informação (ou seja, talvez ele espere por muito tempo na linha durante a sua terceira ou quarta chamada). Essa reincidência (3 ou 4 tentativas) aumenta a taxa de abandono e ainda degrada o nível de serviço, já que, na sua última tentativa, ele provavelmente vai aguardar mais do que 10 segundos para ser atendido. E, como se isso não bastasse, essas tentativas ainda aumentam o

volume de ligações (aparentemente sem explicação, quando, na verdade, trata-se do mesmo cliente ligando várias vezes), mascarando a causa e prejudicando as futuras previsões de demanda.

Todo este processo de dimensionamento e escalonamento dos agentes é realizado inicialmente de forma separada – para cada um dos produtos (básico e *plus*), mas não para cada uma das centrais (Bahia e Pernambuco) – para depois ser consolidado. Uma espécie de "curva média" das duas centrais é utilizada para o dimensionamento de cada produto.

É imprescindível fazer esse processo de forma isolada para cada produto porque é necessário priorizar o cliente *plus*, que precisa ser atendido com um nível de serviço um pouco (mas não muito) melhor (ou seja, mais rapidamente) do que o do cliente básico, já que a Contax negocia um preço diferenciado para a prestação do serviço para aquele tipo de cliente. Durante o processo, o ideal é tentar manter o nível de serviço do cliente básico na faixa de 75% e o do cliente na *plus* na faixa de 85%. Então, é natural tentar melhorar um pouco o escalonamento dos agentes *plus* em detrimento da escala dos básicos, sem, no entanto, comprometer a taxa de abandono acordada no SLA (2%) para o cliente básico.

Assim, depois de fazer toda a distribuição de recursos para o produto básico, espalhar o excesso de atendentes ao longo do dia e analisar a coerência do resultado, é hora do analista verificar o outro produto – o *plus* – e, no caso de haver um déficit de agentes, eventualmente programar uma transferência de recursos oriundos do produto básico.

Já em relação às centrais, é importante realizar o dimensionamento de forma consolidada por causa da possibilidade de fazer transbordo de uma central para a outra. Este transbordo é vantajoso porque existem mais operadores na Bahia, central cuja curva de comportamento intradiário da demanda sofre uma queda mais acentuada na parte da tarde (quando comparada à queda vespertina da central de Pernambuco). Dessa forma, é possível transferir uma parte dos atendentes da Bahia para Pernambuco na parte da tarde.

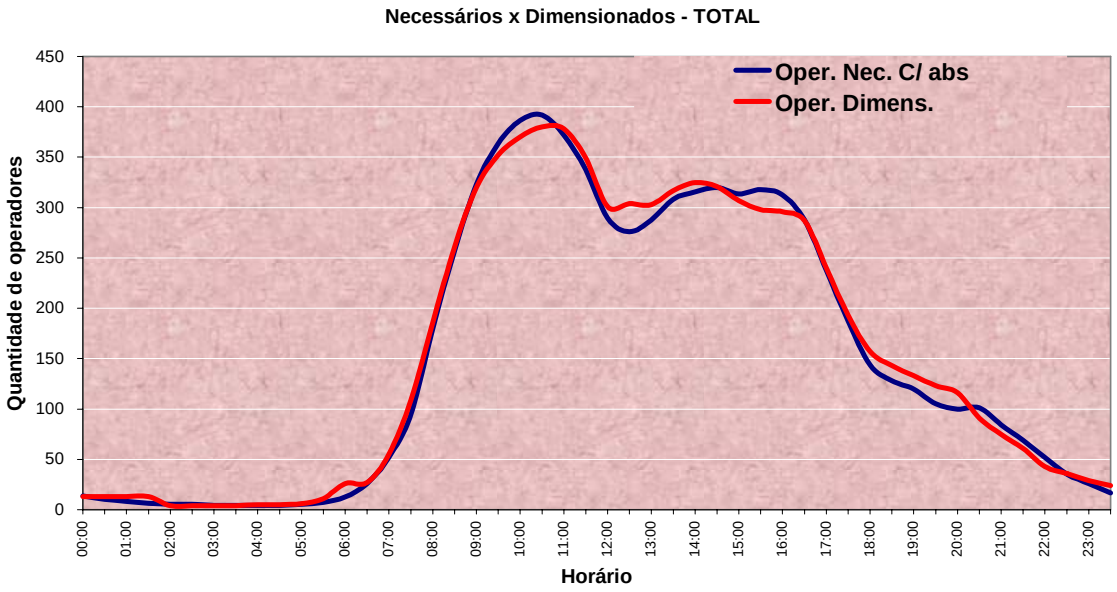
Se o dimensionamento fosse feito de forma isolada para cada central, teria que ser programado um escalonamento de mais agentes (mais custos) em Pernambuco para atender a demanda da tarde, enquanto parte dos operadores baianos estaria ociosa durante o mesmo período. Com o transbordo, é possível manter a quantidade de atendentes de Pernambuco em um nível mais baixo e pedir ajuda à mão-de-obra da Bahia no período vespertino.

O último passo da análise intradiária consiste na programação de horas-extras e na utilização do banco de horas. Depois de realizada a distribuição de operadores sem a utilização destes recursos, é possível verificar em quais horários do dia mão-de-obra adicional está sendo necessária. Se houver disponibilidade de banco de horas, ele pode ser usado; caso contrário, é preciso utilizar horas-extras. Se não houver déficit de atendentes, o banco de horas (quando disponível) pode ser guardado para uma ocasião posterior.

A maior parte dos funcionários prefere realizar horas-extras e/ou compensar as horas do banco na parte da manhã, antes de iniciar a sua jornada. A adesão é maior neste período do dia do que à noite (depois da jornada) mesmo havendo menos operadores entrando no fim da manhã e início da tarde do que saindo no final da tarde e início da noite.

Com a aplicação destes recursos da melhor forma possível, os níveis de serviço desejados para cada produto podem ser agora atingidos. E, então, passa a ser possível observar o resultado do esforço de dimensionamento e escalonamento. As figuras 20 (anterior) e 23 (a seguir) ilustram a relação entre a capacidade demandada e a disponível, ao longo do dia, em dois momentos distintos: antes do dimensionamento (com a configuração de atendentes do mês anterior, recebida da operação) e após a otimização dos recursos, respectivamente.

Figura 23 – Necessidade e disponibilidade (após o escalonamento) de agentes por período, produto 102 (*plus*), Brasil, agosto de 2006



Fonte: Material coletado nas entrevistas

Depois de tantas modificações na busca do melhor dimensionamento, é interessante fazer o registro de todas elas em um resumo a ser enviado para a operação, juntamente com as planilhas contendo o escalonamento sugerido para o mês subseqüente. A visualização das mudanças realizadas fica mais fácil através da indicação via setas – na própria planilha, conforme ilustrado na figura 24 a seguir – das movimentações que foram feitas: de onde foram retirados e para onde foram os atendentes realocados, mudanças de horário, alteração do lanche, migração de básico para *plus*, etc..

Figura 24– Mudanças realizadas durante o escalonamento dos atendentes (6 horas), produto 102 (*plus*), Brasil, agosto de 2006

Horário	06:00	06:30	07:00	07:30	08:00	08:30	09:00	09:30	10:00	10:30
02:30										
03:00										
03:30										
04:00										
04:30										
05:00										
05:30										
06:00	20									
06:30	20	1								
07:00	20	1	21							
07:30	20	1	21	25						
08:00	15	1	21	25	72					
08:30	17	0	15	25	72	40				
09:00	18	1	20	21	72	40	40			
09:30	20	1	21	24	72	40	40	26		
10:00	20	1	21	24	72	40	40	26	18	
10:30	20	1	21	25	72	40	40	26	18	9

Fonte: Material coletado nas entrevistas

As quantidades de atendentes disponíveis destacadas em azul são menores do que as de operadores que iniciam a jornada em cada período por causa da ocorrência do horário de lanche para parte dos funcionários nestes momentos.

Ao receber o resumo do efetivo (ilustrado na figura 25 a seguir) e a planilha informando as modificações sugeridas pela equipe de planejamento, a operação pode eventualmente ajudar – até porque é o setor que melhor conhece o dia-a-dia – criticando algo (através do argumento de que, por exemplo, naquele mês, provavelmente, não haverá adesão para horas-extras no fim do expediente) e sugerindo melhorias (através da citação de que, por exemplo, é preferível, para aquele mês, haver um déficit de agentes ainda maior na parte da manhã do

que um pequeno déficit no início da noite, por ser mais provável conseguir a adesão matutina).

Figura 25 – Resumo do efetivo sugerido pelo dimensionamento para o mês subseqüente, produto 102 (*plus*), Brasil, agosto de 2006

Carga Horária: 06:00h

Horário	Quantidades
0:00 - 6:00	5
5:00 - 11:00	1
5:30 - 11:30	5
6:00 - 12:00	20
6:30 - 12:30	1
7:00 - 13:00	21
7:30 - 13:30	25
8:00 - 14:00	72
8:30 - 14:30	40
9:00 - 15:00	40
9:30 - 15:30	26
10:00 - 16:00	18
10:30 - 16:30	9
11:00 - 17:00	29
11:30 - 17:30	5
12:00 - 18:00	4
12:30 - 18:30	15
13:00 - 19:00	13
13:30 - 19:30	14
14:00 - 20:00	31
14:30 - 20:30	25
15:00 - 21:00	18
15:30 - 21:30	14
16:00 - 22:00	18
16:30 - 22:30	1
17:00 - 23:00	4
17:30 - 23:30	-
18:00 - 0:00	5
18:30 - 0:30	-
19:00 - 1:00	-
19:30 - 1:30	-
20:00 - 2:00	8
Total:	487

Carga Horária: 04:00h

Horário	Quantidades
0:00 - 4:00	-
5:30 - 11:30	-
6:00 - 10:00	-
6:30 - 10:30	-
7:00 - 11:00	7
7:30 - 11:30	32
8:00 - 12:00	8
8:30 - 12:30	39
9:00 - 13:00	16
9:30 - 13:30	-
10:00 - 14:00	-
10:30 - 14:30	-
11:00 - 15:00	-
11:30 - 15:30	-
12:00 - 16:00	-
12:30 - 16:30	-
13:00 - 17:00	12
13:30 - 17:30	31
14:00 - 18:00	50
14:30 - 18:30	10
15:00 - 19:00	-
15:30 - 19:30	-
16:00 - 20:00	-
16:30 - 20:30	-
17:00 - 21:00	-
17:30 - 21:30	-
18:00 - 22:00	-
18:30 - 22:30	5
19:00 - 23:00	3
19:30 - 23:30	4
20:00 - 0:00	13
20:30 - 0:30	-
21:00 - 1:00	-
Total:	230

Fonte: Material coletado nas entrevistas

Em situações como essa, a equipe de planejamento deve refazer o escalonamento e trocar as lacunas, conforme sugerido pela operação. Em outras palavras, existe uma interação entre estes setores e o material enviado anteriormente não consiste, necessariamente, na versão final do plano de trabalho para o próximo mês. A rigor, é possível haver uma imposição por parte da equipe de planejamento, mas o processo não funciona assim, na prática.

O que não pode acontecer é o pessoal da operação executar um plano diferente do que foi sugerido, sem dar *feedback*, criticar nem propor modificações. A equipe de planejamento realizou todos os estudos necessários para gerar o plano proposto e está contando com ele; se não for seguido, sem aviso, pode dar tudo errado na operação.

Enquanto a insubordinação deve ser evitada ao máximo, a crítica construtiva é bem recebida e até incentivada pelos analistas de planejamento, que podem usá-la até para gerar e mudar conceitos acerca do dimensionamento da operação.

Na última etapa do processo de dimensionamento e escalonamento, o analista procura estimar como será o nível de serviço da operação (percentual de chamadas atendidas em até 10 segundos), para todos os dias ao longo do mês (até o momento, os cálculos foram feitos em cima do quinto DMM apenas).

Para tal, a distribuição intradiária de atendentes elaborada durante os passos anteriores (incluindo o uso de horas-extras e banco de horas) é replicada – em uma planilha (ilustrada na tabela 13 a seguir) – para todos os dias do mês, exceto para os sábados e domingos; nesses dias, a programação de agentes é multiplicada por 60% e 40%, respectivamente, já que é essa a proporção de funcionários (dos que trabalham 6 dias por semana) disponíveis em cada um desses dias.

A planilha também recebe como *input* a demanda de ligações prevista para cada um dos dias do mês, assim como os respectivos TMAs. Em cima dessas demandas, são aplicadas as curvas de comportamento intradiário, de acordo com o respectivo dia da semana. Assim, é obtida uma previsão de volume (conforme ilustrado na tabela 13 a seguir) e TMA para cada meia-hora de cada dia do mês em estudo.

Tabela 13 – Escala de atendentes e volume previsto de ligações para cada período de 30 minutos, produto 102 (*plus*), Brasil, novembro de 2005

			VOLUME									
			ter	qua	qui	sex	sáb	dom	seg	ter	qua	qui
			01-nov	02-nov	03-nov	04-nov	05-nov	06-nov	07-nov	08-nov	09-nov	10-nov
ESCALA	0	00:00	0	0	0	0	0	0	0	0	0	0
	0	00:30	0	0	0	0	0	0	0	0	0	0
	0	01:00	0	0	0	0	0	0	0	0	0	0
	0	01:30	0	0	0	0	0	0	0	0	0	0
	0	02:00	0	0	0	0	0	0	0	0	0	0
	0	02:30	0	0	0	0	0	0	0	0	0	0
	0	03:00	0	0	0	0	0	0	0	0	0	0
	0	03:30	0	0	0	0	0	0	0	0	0	0
	0	04:00	0	0	0	0	0	0	0	0	0	0
	0	04:30	0	0	0	0	0	0	0	0	0	0
	0	05:00	0	0	0	0	0	0	0	0	0	0
	0	05:30	0	0	0	0	0	0	0	0	0	0
	10	06:00	323	319	306	286	178	103	345	328	320	300
	20	06:30	776	766	735	687	428	248	828	787	768	721
	36	07:00	1817	1792	1721	1608	1001	580	1939	1841	1797	1687
	78	07:30	3671	3622	3478	3249	2023	1171	3918	3721	3632	3410
	160	08:00	7599	7498	7198	6726	4188	2424	8109	7701	7518	7058
	230	08:30	11398	11245	10797	10088	6282	3636	12162	11551	11277	10586
	290	09:00	14530	14336	13764	12860	8008	4635	15504	14726	14375	13495
	345	09:30	16446	16226	15578	14556	9064	5246	17549	16667	16271	15275
	362	10:00	17727	17489	16791	15689	9769	5655	18915	17965	17538	16464
	362	10:30	18452	18205	17478	16331	10169	5886	19689	18700	18255	17138
	340	11:00	17461	17227	16539	15454	9623	5570	18631	17695	17275	16217
	340	11:30	16230	16013	15374	14365	8945	5177	17318	16448	16057	15074
	276	12:00	13920	13733	13185	12319	7671	4440	14853	14107	13771	12928
	269	12:30	13275	13097	12575	11749	7316	4235	14165	13454	13134	12330
	324	13:00	13451	13271	12742	11905	7413	4291	14353	13632	13308	12493
	313	13:30	13738	13554	13013	12159	7571	4382	14659	13923	13592	12760
	274	14:00	13057	12882	12368	11556	7196	4165	13932	13232	12917	12127
	245	14:30	12754	12583	12081	11288	7029	4068	13609	12925	12618	11845
	241	15:00	12344	12179	11693	10925	6803	3938	13172	12510	12213	11465
	170	15:30	12014	11853	11380	10633	6621	3832	12819	12175	11886	11158
	225	16:00	12196	12032	11552	10794	6721	3890	13013	12360	12066	11327
	215	16:30	11260	11109	10666	9966	6206	3592	12015	11411	11140	10458
	205	17:00	9749	9619	9235	8629	5373	3110	10403	9880	9645	9055
	196	17:30	7935	7829	7517	7023	4373	2531	8467	8042	7851	7370
	195	18:00	6427	6341	6088	5688	3542	2050	6858	6513	6358	5969
	151	18:30	6125	6043	5802	5421	3376	1954	6536	6207	6060	5689
	98	19:00	5233	5163	4957	4631	2884	1669	5583	5303	5177	4860
	97	19:30	4638	4575	4393	4104	2556	1479	4949	4700	4588	4307
	85	20:00	3389	3343	3210	2999	1868	1081	3616	3434	3353	3147
	85	20:30	3265	3221	3093	2889	1799	1041	3484	3309	3230	3032
	45	21:00	2480	2447	2349	2195	1367	791	2647	2514	2454	2304
	45	21:30	2005	1978	1899	1775	1105	640	2140	2032	1984	1862
	0	22:00	0	0	0	0	0	0	0	0	0	0
	0	22:30	0	0	0	0	0	0	0	0	0	0
	0	23:00	0	0	0	0	0	0	0	0	0	0
	0	23:30	0	0	0	0	0	0	0	0	0	0
	527,3		305.685	301.589	289.557	270.545	168.469	97.513	326.178	309.792	302.428	283.913

Fonte: Material coletado nas entrevistas

A partir do cruzamento das informações de oferta de capacidade e de previsão de demanda (volume e TMA) para cada dia e horário, a planilha estima – de acordo com a mesma metodologia Erlang usada anteriormente – os níveis de serviço que serão obtidos em cada um dos intervalos de tempo, levando em conta as mesmas premissas de absenteísmo e taxa de ocupação aplicadas anteriormente, conforme ilustrado na tabela 14 a seguir.

Tabela 14 – Nível de serviço estimado para cada período de 30 minutos, produto 102 (*plus*), Brasil, novembro de 2005

NÍVEL SERVIÇO										
	ter	qua	qui	sex	sáb	dom	seg	ter	qua	qui
	01-nov	02-nov	03-nov	04-nov	05-nov	06-nov	07-nov	08-nov	09-nov	10-nov
00:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
00:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
01:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
01:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
02:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
02:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
03:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
03:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
04:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
04:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
05:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
05:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
06:00	99%	99%	99%	100%	97%	93%	99%	99%	99%	99%
06:30	99%	100%	100%	100%	99%	99%	99%	99%	100%	100%
07:00	88%	91%	96%	99%	96%	98%	75%	89%	93%	98%
07:30	99%	100%	100%	100%	100%	100%	97%	99%	100%	100%
08:00	100%	100%	100%	100%	100%	100%	99%	100%	100%	100%
08:30	100%	100%	100%	100%	100%	100%	96%	100%	100%	100%
09:00	100%	100%	100%	100%	100%	100%	72%	100%	100%	100%
09:30	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
10:00	100%	100%	100%	100%	100%	100%	99%	100%	100%	100%
10:30	100%	100%	100%	100%	100%	100%	23%	100%	100%	100%
11:00	98%	100%	100%	100%	100%	100%	0%	99%	100%	100%
11:30	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
12:00	100%	100%	100%	100%	100%	100%	95%	100%	100%	100%
12:30	100%	100%	100%	100%	100%	100%	98%	100%	100%	100%
13:00	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
13:30	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
14:00	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
14:30	94%	99%	100%	100%	100%	100%	0%	96%	100%	100%
15:00	99%	100%	100%	100%	100%	100%	67%	100%	100%	100%
15:30	0%	0%	0%	0%	0%	86%	0%	0%	0%	0%
16:00	25%	83%	100%	100%	100%	100%	0%	47%	92%	100%
16:30	95%	99%	100%	100%	100%	100%	0%	96%	99%	100%
17:00	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
17:30	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
18:00	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
18:30	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
19:00	40%	70%	96%	100%	99%	100%	0%	50%	80%	99%
19:30	99%	100%	100%	100%	100%	100%	96%	99%	100%	100%
20:00	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20:30	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
21:00	0%	5%	64%	92%	85%	95%	0%	0%	22%	79%
21:30	100%	100%	100%	100%	100%	100%	99%	100%	100%	100%
22:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
22:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
23:00	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
23:30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	91%	94%	96%	96%	96%	99%	68%	92%	95%	96%

Fonte: Material coletado nas entrevistas

Finalmente, é elaborado um resumo (exemplificado na tabela 15 a seguir) relacionando o volume, o TMA e o nível de serviço estimados para cada um dos dias do mês em análise, onde estes são hierarquizados de acordo com os seus volumes (DMM1, DMM2, ... , DMM30). Assim, é possível identificar os dias em que o nível de serviço se situará abaixo da meta (os dias de maior movimento).

Tabela 15 – Volume, TMA e nível de serviço estimados para os dias (hierarquizados) de novembro de 2005, produto 102 (*plus*), Brasil

Dia	Volume	TMA	NS	DMM
1-nov	305.685	31	91%	DMM 8
2-nov	301.589	31	94%	DMM 10
3-nov	289.557	31	96%	DMM 13
4-nov	270.545	31	96%	DMM 16
5-nov	168.469	29	96%	DMM 24
6-nov	97.513	28	99%	DMM 30
7-nov	326.178	30	68%	DMM 3
8-nov	309.792	30	92%	DMM 9
9-nov	302.428	30	95%	DMM 11
10-nov	283.913	30	96%	DMM 14
11-nov	266.151	30	96%	DMM 17
12-nov	161.464	29	96%	DMM 26
13-nov	101.564	28	99%	DMM 29
14-nov	319.115	30	85%	DMM 5
15-nov	285.593	26	95%	DMM 21
16-nov	304.277	25	96%	DMM 19
17-nov	304.941	25	96%	DMM 20
18-nov	266.208	29	96%	DMM 18
19-nov	160.696	29	96%	DMM 25
20-nov	101.019	28	98%	DMM 28
21-nov	325.122	30	75%	DMM 4
22-nov	315.182	30	89%	DMM 7
23-nov	331.284	30	55%	DMM 2
24-nov	293.243	31	95%	DMM 12
25-nov	274.557	31	96%	DMM 15
26-nov	167.871	29	96%	DMM 23
27-nov	104.273	28	96%	DMM 27
28-nov	344.253	30	38%	DMM 1
29-nov	201.680	30	100%	DMM 22
30-nov	319.043	30	86%	DMM 6
	7.685.492	**30**	**88,1%**	

Fonte: Material coletado nas entrevistas

Depois de validada, esta planilha que estima o nível de serviço ao longo de todo o mês também é enviada para a operação. Ela servirá de base para o acompanhamento do cotidiano operacional e para pequenos ajustes diários, depois que a operação implementar na prática o seu conteúdo.

Infelizmente, a equipe de planejamento não consegue mensurar – com um nível de detalhe muito grande – se o nível de serviço efetivamente ocorrido foi similar ao planejado. Para tal

mensuração poder ser bem feita, as premissas estimadas (da demanda e do próprio dimensionamento) teriam que acontecer com bastante acurácia, na realidade.

Mas o volume de demanda real não necessariamente bate com o que foi previsto. Além disso, nem sempre o escalonamento elaborado de agentes (quantidades certas em cada horário) é cumprido, já que, entre outros motivos, o banco de horas nem sempre pode ser utilizado da forma planejado e o absenteísmo, às vezes, acontece com maior ou menor intensidade do que consta na premissa.

A demanda real até que coincide – com alguma freqüência – com a que foi prevista, mas a correta execução do escalonamento planejado não é tão comum assim de acontecer. Para uma correta mensuração acerca da acurácia da previsão do nível de serviço, teria que ser feita uma grande pesquisa para ser estudado um dia em que a demanda real foi muito próxima da prevista e as escalas ocorreram exatamente como foram dimensionadas, o que é difícil de acontecer.

Por isso, o setor não trabalha com nenhuma planilha que sumarize essa comparação. O nível de serviço é acompanhado mesmo diariamente, mais através do "sentimento" e da percepção da equipe de planejamento do que propriamente de um procedimento formal. Em outras palavras, uma análise macro é realizada a esse respeito.

Por exemplo, se em um determinado dia, o nível de serviço previsto era de 90% (número fechado para o dia, e não aberto por horário) e o obtido foi 89% (ou 88% ou 91% ou 92%), isso é considerado um acerto, já que a previsão tem margem de erro para cima e para baixo. Agora, se nesse dia o nível de serviço obtido for 60%, é necessário checar se o volume e TMA reais fugiram muito dos previstos (nesse caso, a "culpa" é da previsão de demanda); caso contrário, o absenteísmo efetivo precisa ser comparado com o que foi utilizado nas premissas. Caso não seja detectado nenhum grande desvio, é preciso fazer uma análise micro para detectar em qual(is) horário(s) houve problema e perda de nível de serviço.

Nesse caso, é preciso entender o motivo de ter ocorrido um desvio exagerado no final do dia. Uma possível explicação é a chamada "inversão da curva", que ocorre quando a maior demanda for prevista para a manhã (gerando um dimensionamento mais numeroso para essa parte do dia e em menor escala para a tarde) e o inverso ocorre: maior demanda no momento (tarde) de menor capacidade de atendimento, acarretando um excelente nível de serviço

(100%, com ociosidade, provavelmente) pela manhã, mas uma grande deterioração no mesmo na parte da tarde. Em uma situação como essas, a análise macro não é suficiente, pois o volume diário de demanda pode ter se comportado conforme previsto, variando apenas a sua oscilação intradiária.

Mas, se a curva intradiária tiver se comportado conforme foi planejado, pode ter havido um erro sério no processo de dimensionamento. No entanto, raramente a análise chega a esse nível de detalhe; é muito pouco comum acontecer de o volume, o TMA e a quantidade de pessoas estarem coerentes e o nível de serviço fugir muito do estimado.

Na opinião da equipe de planejamento, a fórmula Erlang utilizada para o cálculo do nível de serviço não é totalmente precisa, mas não é radicalmente imprecisa a ponto de chamar a atenção e gerar situações onde as premissas sejam respeitadas e o nível de serviço se distancie muito do calculado.

Na verdade, a equipe questiona um pouco a fórmula de cálculo do nível de serviço em função da distribuição de operadores, através das fórmulas Erlang. Esse questionamento nasceu de algumas observações empíricas, como em uma situação em que o nível de serviço calculado para um horário com déficit de 3 agentes era 77% e passava para 0% quando o déficit subia para 12 agentes.

O sentimento geral é de que essa forma de cálculo não é 100% precisa, mas gera resultados próximos da realidade, já que entram fatores como tempo de espera, entre outros. Não existe concordância de que essa é a maneira ideal de calcular o nível de serviço, mas o setor não encontrou nenhuma outra abordagem mais precisa, em pesquisas coletando informações e planilhas verificadas.

6. Análise do Caso

Com o intuito de responder as questões propostas na seção 1.1.2 de forma a atingir os objetivos estipulados na seção 1.1.1, este capítulo procura atacar – através de metodologias estudadas e sugeridas pelo pesquisador – os problemas levantados na seção 5.2: a previsão de demanda e o dimensionamento da capacidade de atendimento.

6.1. Previsão de demanda

Hoje em dia, conforme descrito na seção 5.2.1, a previsão básica do produto 103 é extraída do TotalView e a equipe de previsão realiza manualmente – no Excel – a inserção das informações a respeito da postagem e vencimento das contas dos clientes, dentro de cada ciclo. Essas informações são fornecidas pelo setor de faturamento e são capazes de impactar a demanda de ligações em cada dia, assim como o tempo médio de atendimento (TMA). A reivindicação geral consiste na criação de uma forma através da qual as informações do faturamento pudessem ser usadas de uma maneira mais sistemática no processo de elaboração das previsões. A idéia é evitar o trabalho manual e o uso apenas da intuição no analista, reduzindo a possibilidade de erro.

O objetivo torna-se, então, construir uma ferramenta de previsão capaz de capturar de forma metódica as informações relevantes dos ciclos de contas – fugindo do sentimento, da intuição e do trabalho manual dos analistas – e de conseguir gerar resultados mais acurados. A idéia é a ferramenta não substituir a intuição e experiência da equipe de previsão, mas servir de apoio à sua tomada de decisão.

A equipe responsável pela previsão de demanda (em especial o Gerente de Planejamento de Tráfego) deu bastante apoio à idéia de implantação de uma nova ferramenta capaz de facilitar e melhorar a qualidade do processo e a atividade de cada membro da equipe.

Até por isso, a coleta de informações necessárias para o desenvolvimento da ferramenta fluiu com relativa facilidade e sem grande resistência, a menos de atrasos (até de algumas semanas) ocorridos durante o envio do material que, inclusive, inviabilizaram a mensuração da funcionalidade do modelo em relação ao "futuro", conforme foi explicado na seção 4.4. Acredita-se, no entanto, que estes atrasos tenham sido motivados apenas por problemas de

saúde e de excesso de trabalho do responsável, e não por uma eventual resistência do mesmo à utilização de uma nova abordagem para a atividade de previsão de demanda.

Vale relatar que a real magnitude dos valores informados foi camuflada (através do uso de uma constante de proporcionalidade) pela necessidade de sigilo da informação (e não motivada por algum outro fator), conforme será detalhado na seção 6.1.1 a seguir.

Adicionalmente – talvez em parte motivada pela necessidade de valorização do próprio trabalho, mas certamente apresentando coerência – uma ressalva generalizada foi levantada pela equipe, alertando para a necessidade de a nova ferramenta não impedir nem engessar as contribuições razoavelmente subjetivas (e baseadas na experiência) que os responsáveis pela previsão costumam agregar – de forma positiva – às predições preliminares sugeridas pelo sistema.

6.1.1. A escolha do método de previsão e do "produto-cobaia"

Operacionalmente, a equipe de previsão coletaria, com a antecedência usual, os *inputs* necessários (oriundos do setor de faturamento) e alimentaria a ferramenta de previsão – ambientada em uma planilha de Excel – com eles. O *output* da ferramenta consistiria na quantidade de ligações demandada e o tempo médio de atendimento para cada data futura.

O 103 é o produto ideal para a construção de tal ferramenta, principalmente por dois motivos:

- é o produto que parece encerrar o maior potencial de melhoria na sua previsão de demanda, já que a empresa dispõe de informações (quantidades de contas enviadas para os clientes em cada data, assim como suas datas de vencimento) que poderiam fornecer uma melhor acurácia, estando apenas carente de um método mais adequado;

- a demanda do produto, segundo a Coordenação de Planejamento de Tráfego, apresenta pouca sazonalidade ao longo do ano – e, como o histórico de dados fornecido pela empresa para a construção da ferramenta abrange um período de cerca de 1 ano, a sazonalidade mensal não poderia ser capturada e isolada, o que acabaria por poluir o processo de previsão de um produto cuja demanda fosse sazonal.

O método escolhido foi o da Regressão Múltipla, pois é capaz de capturar o impacto específico de cada evento e informação na demanda e no TMA de cada dia. Por exemplo, qual o impacto – na demanda de ligações – do fato do dia em questão cair em uma quinta-feira?

Qual o impacto – no TMA – do dia consistir em um feriado? Qual o impacto – nas duas grandezas – de chegarem 200 mil contas para os clientes no dia em questão? E qual o impacto de vencerem 300 mil contas neste dia? E do dia ser a véspera do vencimento de 500 mil contas?

A idéia é verificar a distância do dia específico em relação aos eventos especiais (chegada e vencimento das contas) e calcular para quantas contas esse dia é o **C** (dia de chegada), o **V** (dia do vencimento), o **V – 1** (véspera do vencimento), o **V – 2** (antevéspera do vencimento), ... , **V + 1** (dia seguinte ao vencimento), **V + 2**, ...

De posse do cronograma de envio e vencimento de contas do faturamento (fornecido pela equipe de previsão, para o período de julho de 2005 a julho de 2006), apresentado parcialmente na tabela 6, na seção 5.2.1, foi possível prever a data de chegada para cada um dos ciclos de contas: de uma forma geral, a chegada ocorre um dia após a postagem, exceto para a Baixada Fluminense e o interior do Estado do RJ (ciclo RJ-7), cuja chegada das contas acontece dois dias após a postagem.

Depois disso, os ciclos de contas foram consolidados a um nível nacional – até porque, conforme descrito na seção 5.2.1, tudo funciona, em termos organizacionais, como se houvesse uma central única de atendimento, abrangendo potencialmente os clientes de todos os estados – em termos da data de chegada e data de vencimento das contas, conforme apresentado parcialmente nas tabelas 16 e 17, respectivamente e a seguir.

Tabela 16 – Total de contas agrupadas por data de chegada prevista, setembro a outubro de 2005

Soma de Contas	
Data de Chegada Prevista	Total
06/09/05	3.360.569
13/09/05	801.143
14/09/05	432.670
15/09/05	709.437
30/08/05	2.181.734
17/08/05	246.901
26/09/05	1.315.342
29/09/05	2.506.038
04/10/05	1.808.735
05/10/05	1.167.447
11/10/05	1.890.364
27/09/05	820.855
28/09/05	1.555.097
07/10/05	800.105
18/10/05	709.437
26/10/05	1.731.518
28/10/05	587.118
Total geral	133.108.943

Fonte: Planilha elaborada pelo pesquisador

Tabela 17 – Total de contas agrupadas por data de vencimento, setembro a outubro de 2005

Soma de Contas	
Vencimento	Total
04/09/05	2.142.215
07/09/05	1.918.920
11/09/05	2.177.552
13/09/05	798.630
15/09/05	1.456.656
17/09/05	801.143
20/09/05	432.670
23/09/05	709.437
23/08/05	246.901
01/10/05	2.136.197
04/10/05	2.142.215
07/10/05	1.918.920
11/10/05	2.177.552
13/10/05	798.630
15/10/05	1.456.656
17/10/05	801.143
20/10/05	432.670
23/10/05	709.437
Total geral	133.108.943

Fonte: Planilha elaborada pelo pesquisador

Para poder ser desenvolvido o modelo de regressão múltipla para prever a quantidade de ligações e o TMA em cada data, foi preciso ter acesso ao histórico destas duas grandezas. Por motivo de sigilo estratégico, esses valores não puderam ser revelados, em sua plena magnitude; no entanto, a eles foi aplicada uma constante de proporcionalidade, e os resultados, de julho de 2005 até março de 2006, foram fornecidos pela empresa e estão apresentados parcialmente na tabela 18 a seguir.

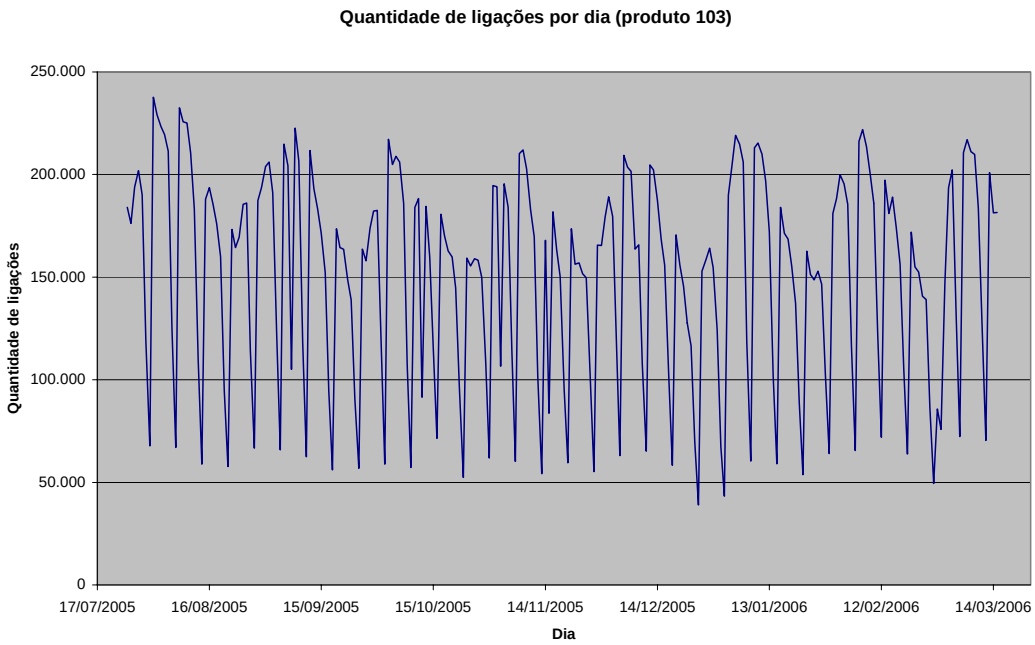
Tabela 18 – Histórico da quantidade de ligações e do TMA (multiplicados por uma constante) para o produto 103, novembro de 2005

Data	Ligações	TMA (seg)
01/11/2005	193.977	145
02/11/2005	106.587	117
03/11/2005	195.434	146
04/11/2005	183.938	149
05/11/2005	114.623	135
06/11/2005	60.267	121
07/11/2005	210.236	150
08/11/2005	211.949	153
09/11/2005	202.792	154
10/11/2005	182.913	152
11/11/2005	169.596	146
12/11/2005	101.658	134
13/11/2005	54.152	115
14/11/2005	167.911	140
15/11/2005	83.714	113
16/11/2005	181.821	143
17/11/2005	163.432	146
18/11/2005	150.299	138
19/11/2005	95.204	123
20/11/2005	59.552	92
21/11/2005	173.555	146
22/11/2005	156.347	143
23/11/2005	156.896	143
24/11/2005	151.793	144
25/11/2005	149.621	137
26/11/2005	100.965	124
27/11/2005	55.197	110
28/11/2005	165.545	138
29/11/2005	165.424	138
30/11/2005	179.283	139

Fonte: Material coletado nas entrevistas

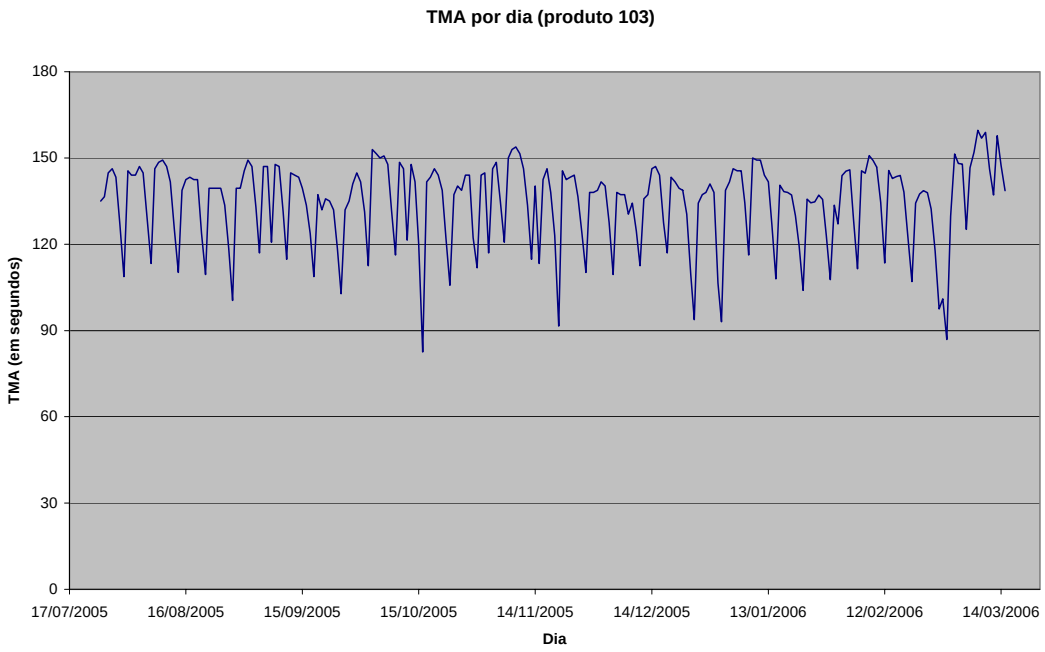
O comportamento das duas grandezas em função do tempo pode ser mais bem observado através dos gráficos apresentados a seguir, nas figuras 26 e 27, respectivamente.

Figura 26 – Quantidade de ligações por dia (multiplicada por uma constante), para o produto 103, julho de 2005 a março de 2006



Fonte: Gráfico elaborado pelo pesquisador

Figura 27 – TMA por dia (multiplicado por uma constante), para o produto 103, julho de 2005 a março de 2006



Fonte: Gráfico elaborado pelo pesquisador

A série da quantidade de ligações não apresenta nenhuma tendência de crescimento ou decrescimento (apenas um comportamento errático), dispensando, assim, a necessidade de inclusão de uma variável explicativa para registrar a passagem do tempo (o que seria obrigatório caso a grandeza mostrasse um crescimento ou decrescimento sistemático).

Assim como a quantidade de ligações, o tempo médio de atendimento não apresenta tendência alguma, necessitando do mesmo tratamento; no entanto, observa-se uma diferença significativa em relação à grandeza anterior: o TMA é razoavelmente mais homogêneo que a quantidade de ligações, o que torna a sua magnitude um pouco mais previsível.

6.1.2. O modelo de previsão para a demanda de ligações

De posse de todas essas informações foi possível montar a planilha de Excel com os dados necessários para a construção do modelo de regressão múltipla dentro do qual o comportamento da variável dependente "quantidade de ligações em um dia específico" – inicialmente – poderia estar sendo explicado por 10 variáveis independentes:

- dia da semana;
- feriado – se o dia específico consiste em um feriado ou não;
- chegada (C) – quantas contas estão previstas para chegarem ao cliente no dia específico;
- vencimento (V) – quantas contas vencem no dia específico;
- V - 3 – para quantas contas aquele dia específico corresponde a 3 dias antes do vencimento, ou, em outras palavras, quantas contas irão vencer 3 dias depois da data específica;
- V - 2 – para quantas contas aquele dia específico corresponde à antevéspera do vencimento;
- V - 1 – para quantas contas aquele dia específico corresponde à véspera do vencimento;

- V + 1 – para quantas contas aquele dia específico corresponde a 1 dia depois do vencimento, ou, em outras palavras, quantas contas terão vencido na véspera daquela data específica;

- V + 2 – para quantas contas aquele dia específico corresponde a 2 dias depois do vencimento;

- V + 3 – para quantas contas aquele dia específico corresponde a 3 dias depois do vencimento.

A tabela 19 a seguir apresenta parcialmente a organização da planilha de dados em questão.

Tabela 19 – Planilha de dados do modelo de regressão múltipla para prever a quantidade de ligações, produto 103, fevereiro de 2006

Dia	Dia da semana	Feriado ?	Chegada (C)	V - 3	V - 2	V - 1	Vencimento (V)	V + 1	V + 2	V + 3	Ligações
01/02/2006	4a. feira	Não	2.506.038	2.142.215	-	-	2.136.197	-	-	-	199.908
02/02/2006	5a. feira	Não	-	-	2.142.215	-	-	2.136.197	-	-	195.497
03/02/2006	6a. feira	Não	1.122.892	-	-	2.142.215	-	-	2.136.197	-	185.542
04/02/2006	Sábado	Não	-	1.918.920	-	-	2.142.215	-	-	2.136.197	116.657
05/02/2006	Domingo	Não	-	-	1.918.920	-	-	2.142.215	-	-	65.558
06/02/2006	2a. feira	Não	587.354	-	-	1.918.920	-	-	2.142.215	-	216.104
07/02/2006	3a. feira	Não	1.484.473	-	-	-	1.918.920	-	-	2.142.215	221.831
08/02/2006	4a. feira	Não	1.238.119	2.177.552	-	-	-	1.918.920	-	-	213.584
09/02/2006	5a. feira	Não	801.143	-	2.177.552	-	-	-	1.918.920	-	200.359
10/02/2006	6a. feira	Não	-	798.630	-	2.177.552	-	-	-	1.918.920	185.882
11/02/2006	Sábado	Não	-	-	798.630	-	2.177.552	-	-	-	122.764
12/02/2006	Domingo	Não	-	1.456.656	-	798.630	-	2.177.552	-	-	72.033
13/02/2006	2a. feira	Não	-	-	1.456.656	-	798.630	-	2.177.552	-	197.231
14/02/2006	3a. feira	Não	432.670	801.143	-	1.456.656	-	798.630	-	2.177.552	181.092
15/02/2006	4a. feira	Não	709.437	-	801.143	-	1.456.656	-	798.630	-	188.990
16/02/2006	5a. feira	Não	-	-	-	801.143	-	1.456.656	-	798.630	174.253
17/02/2006	6a. feira	Não	-	432.670	-	-	801.143	-	1.456.656	-	156.317
18/02/2006	Sábado	Não	-	-	432.670	-	-	801.143	-	1.456.656	105.306
19/02/2006	Domingo	Não	-	-	-	432.670	-	-	801.143	-	63.827
20/02/2006	2a. feira	Não	-	709.437	-	-	432.670	-	-	801.143	171.921
21/02/2006	3a. feira	Não	-	-	709.437	-	-	432.670	-	-	155.035
22/02/2006	4a. feira	Não	2.136.197	-	-	709.437	-	-	432.670	-	152.537
23/02/2006	5a. feira	Não	1.555.097	-	-	-	709.437	-	-	432.670	140.621
24/02/2006	6a. feira	Não	2.506.038	-	-	-	-	709.437	-	-	139.056
25/02/2006	Sábado	Não	-	-	-	-	-	-	709.437	-	86.746
26/02/2006	Domingo	Não	-	2.136.197	-	-	-	-	-	709.437	49.520
27/02/2006	2a. feira	Não	-	-	2.136.197	-	-	-	-	-	85.790
28/02/2006	3a. feira	Sim	-	-	-	2.136.197	-	-	-	-	75.801

Fonte: Planilha elaborada pelo pesquisador

Cabe salientar que as variáveis independentes "dia da semana" e "feriado" foram trabalhadas como sendo do tipo *dummy* (ou simbólicas ou binárias). Assim, e de acordo com a revisão bibliográfica apresentada na seção 3.1.2.1, foram criadas 6 variáveis *dummy* para o dia da

semana (**domingo**, **segunda-feira**, **terça-feira**, **quarta-feira**, **quinta-feira** e **sexta-feira**) e 1 variável *dummy* para a ocorrência do feriado (**feriado**).

Dessa forma, para o dia 28/02/2006, por exemplo, a sua variável *dummy* **feriado** assumiu o valor 1 (0 = Não; 1 = Sim), por se tratar da terça-feira de Carnaval; e as suas variáveis *dummy* **domingo**, **segunda-feira**, **quarta-feira**, **quinta-feira** e **sexta-feira** assumiram o valor 0 enquanto que a sua variável dummy **terça-feira** assumiu o valor 1. Seguindo o mesmo raciocínio, para o dia 04/02/2006, por exemplo, a sua variável *dummy* **feriado** assumiu o valor 0; e as suas variáveis *dummy* **domingo**, **segunda-feira**, **terça-feira**, **quarta-feira**, **quinta-feira** e **sexta-feira** assumiram o valor 0, caracterizando, portanto e por exclusão, o dia em questão como sendo um sábado (seria redundante a existência de uma variável *dummy* **sábado**, pois este dia já estará sendo caracterizado quando todas as variáveis *dummy* relativas ao dia da semana assumirem o valor 0).

Isto posto, foi utilizada a ferramenta de "Regressão" do suplemento de "Análise de Dados" do Excel para gerar o modelo de regressão múltipla proposto. As tabelas 20, 21 e 22 a seguir foram extraídas do relatório fornecido pela ferramenta em questão.

Tabela 20 – Estatística de regressão do modelo de regressão múltipla para prever a quantidade de ligações, produto 103

Estatística de regressão	
R múltiplo	0,92
R-Quadrado	0,85
R-quadrado ajustado	**0,84**
Erro padrão	20.320
Observações	234

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

A primeira análise aponta para um modelo com um potencial muito bom, já que 84% (R-quadrado ajustado) da variação dos dados pode ser explicada por ele. Mas é fundamental verificar também o F de significação, destacado na tabela 21 a seguir.

Tabela 21 – Análise de Variância do modelo de regressão múltipla para prever a quantidade de ligações, produto 103

ANOVA

	gl	*F*	*F de significação*
Regressão	15,00	85,64	**0,0000**
Resíduo	218,00		
Total	233,00		

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

O F de significação resultante foi 0,0000 (zero). Neste caso, pode-se acreditar na real influência das variáveis explicativas na quantidade de ligações e aceitar o modelo com bastante certeza.

Para analisar a contribuição e a representatividade de cada variável independente, é necessário examinar seus coeficientes, juntamente com os respectivos *p-value* (ou valor p), apresentados na tabela 22 a seguir.

Tabela 22 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo de regressão múltipla para prever a quantidade de ligações, produto 103

	Coeficientes	*valor-P*	*95% inferiores*	*95% superiores*
Interseção	**92.545**	0,00%	84.600	100.490
Domingo	**(41.228)**	**0,00%**	(51.225)	(31.232)
2a. Feira	**79.882**	**0,00%**	69.994	89.769
3a. Feira	**75.694**	**0,00%**	65.044	86.343
4a. Feira	**73.812**	**0,00%**	62.846	84.779
5a. Feira	**73.068**	**0,00%**	63.079	83.056
6a. Feira	**61.763**	**0,00%**	51.766	71.760
Feriado	**(71.088)**	**0,00%**	(87.237)	(54.938)
Chegada (C)	**0,0035**	**13,57%**	(0,0011)	0,0081
V - 3	**0,0052**	**2,21%**	0,0007	0,0096
V - 2	**0,0042**	**7,81%**	(0,0005)	0,0089
V - 1	**0,0068**	**0,47%**	0,0021	0,0115
Vencimento (V)	**0,0074**	**0,19%**	0,0027	0,0120
V + 1	**0,0071**	**0,30%**	0,0024	0,0117
V + 2	**0,0064**	**0,75%**	0,0017	0,0110
V + 3	**0,0009**	**68,36%**	(0,0035)	0,0053

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

Ao analisar a tabela 22 anterior, é possível perceber um altíssimo valor p para a variável V + 3, indicando que a "quantidade de contas para as quais aquele dia específico corresponde a 3 dias depois do vencimento" não está funcionando como um bom explicador linear para a quantidade de ligações.

O efeito "pós-vencimento" parece, então, acabar no segundo dia após o vencimento da conta (já que os valores p de V + 1 e V + 2 são bem baixos); em outras palavras, os clientes ainda ligam para o 103 um ou dois dias após o vencimento da conta, mas esse impacto cai bastante a partir do terceiro dia, deixando de ser significativo.

Esta variável (V + 3) pode estar "poluindo" o modelo e convém retirá-la da análise. A ferramenta de "Regressão" do Excel foi novamente usada para gerar o modelo de regressão múltipla. As tabelas 23, 24 e 25 a seguir foram extraídas do relatório fornecido pela ferramenta.

Tabela 23 – Estatística de regressão do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103

Estatística de regressão	
R múltiplo	0,92
R-Quadrado	0,85
R-quadrado ajustado	**0,85**
Erro padrão	20.282
Observações	234

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

O modelo continuou com um potencial muito bom (até ligeiramente melhor), já que 85% (R-quadrado ajustado) da variação dos dados pode ser explicada por ele. O F de significação pode ser observado na tabela 24 a seguir.

O F de significação resultante foi novamente 0,0000 (zero), acarretando na aceitação do modelo com bastante certeza.

Tabela 24 – Análise de Variância do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103

ANOVA

	gl	*F*	*F de significação*
Regressão	14,00	92,09	**0,0000**
Resíduo	219,00		
Total	233,00		

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

Os coeficientes linear e angulares, juntamente com os respectivos valores p, estão apresentados na tabela 25 a seguir.

Tabela 25 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103

	Coeficientes	*valor-P*	*95% inferiores*	*95% superiores*
Interseção	**92.693**	0,00%	84.795	100.591
Domingo	**(41.103)**	**0,00%**	(51.062)	(31.144)
2a. Feira	**79.874**	**0,00%**	70.005	89.742
3a. Feira	**75.900**	**0,00%**	65.318	86.483
4a. Feira	**73.908**	**0,00%**	62.972	84.844
5a. Feira	**73.082**	**0,00%**	63.113	83.051
6a. Feira	**61.898**	**0,00%**	51.941	71.855
Feriado	**(71.345)**	**0,00%**	(87.416)	(55.275)
Chegada (C)	**0,0034**	**14,30%**	(0,0012)	0,0080
V - 3	**0,0052**	**2,11%**	0,0008	0,0096
V - 2	**0,0042**	**7,86%**	(0,0005)	0,0089
V - 1	**0,0072**	**0,13%**	0,0028	0,0115
Vencimento (V)	**0,0077**	**0,06%**	0,0034	0,0120
V + 1	**0,0071**	**0,30%**	0,0024	0,0117
V + 2	**0,0060**	**0,71%**	0,0017	0,0104

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

Como pode ser observado, os valores p das outras variáveis mudaram muito pouco e, para este modelo modificado, apenas duas variáveis apresentam coeficientes angulares com valores p superiores a 5% (nível de significância mais comumente utilizado nos testes de hipóteses): Chegada (C) e V – 2.

Em relação a esta última, ela será mantida no modelo, porque o valor p do seu coeficiente angular está abaixo de 10% (um valor também bastante utilizado como nível de significância nos testes de hipóteses) e pelo fato das variáveis V – 3 e V – 1 apresentarem-se como bons explicadores lineares (não faria muito sentido manter, no modelo, V – 3 , retirar V – 2 e manter V – 1, pois a continuidade estaria sendo quebrada e seria o mesmo que afirmar que os clientes ligam para o 103 três dias antes do vencimento e na véspera do mesmo, mas não na antevéspera!).

Já a variável Chegada (C) também será mantida no modelo (apesar de apresentar valor p maior do que 10%) pelo fato de ter sido identificada, pela equipe de planejamento, como crítica e muito impactante na demanda de ligações; além disso, seu valor p (14,30%) não ficou tão acima de um nível de significância considerado razoável (10%).

Desta forma, a retirada da variável V + 3 foi benéfica para o modelo (tornando-o, inclusive, mais parcimonioso, isto é, com menos variáveis) e os seus coeficientes podem ser, então, interpretados.

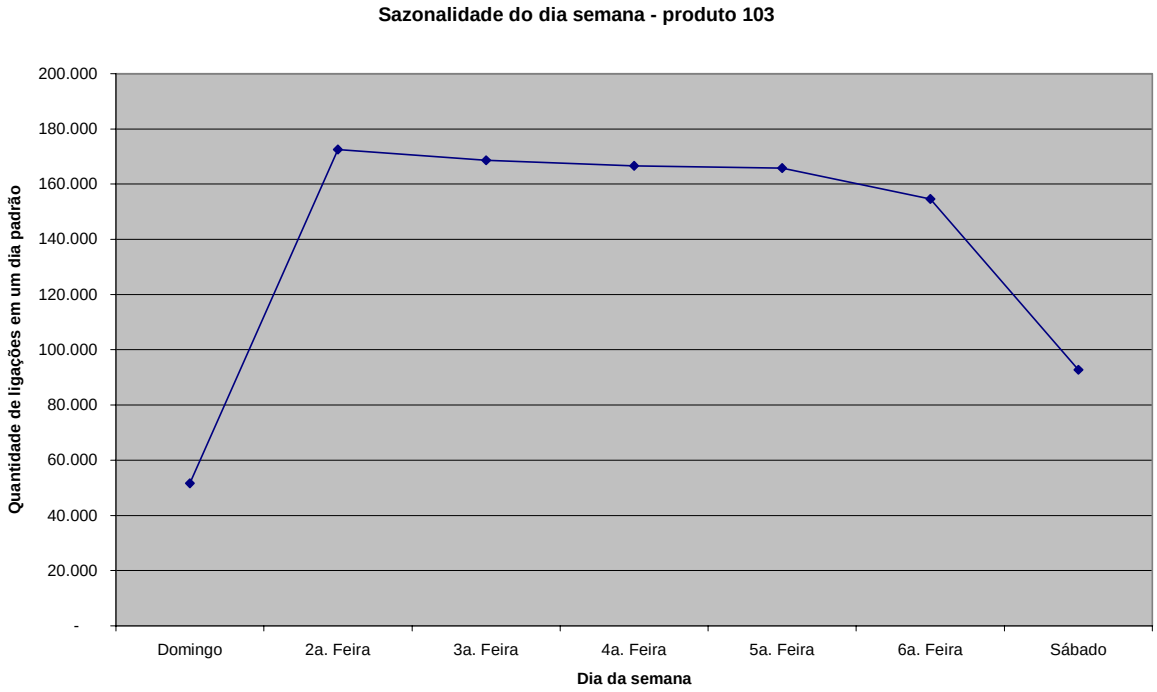
O primeiro coeficiente – rotulado de "Intersecção" – corresponde ao coeficiente linear da regressão, que consiste no valor da variável dependente (no caso, a quantidade de ligações) quando todas as variáveis independentes (inclusive as *dummy*) assumirem o valor zero; ou seja, quando domingo, segunda-feira, terça-feira, quarta-feira, quinta-feira, sexta-feira, feriado, C, V – 3, V – 2, V – 1, V, V + 1 e V + 2 forem zero – ou ainda, quando o dia for sábado, não-feriado e não consistir da data de chegada, nem de nenhuma data dentro do intervalo de 6 dias em torno do vencimento de alguma conta, a previsão é que a quantidade de ligações demandadas seja 92.693.

Esse número pode ser encarado como uma quantidade básica, à qual devem ser adicionados os efeitos da ocorrência das variáveis independentes, expressados através dos seus respectivos coeficientes angulares, destacados na tabela 25 anterior.

Assim, o fato do dia ser domingo contribui com uma diminuição de 41.103 ligações previstas em relação ao número básico (92.693); a segunda-feira espera, segundo o modelo, 79.874 ligações a mais do que o dia básico (sábado); a terça-feira espera 75.900 ligações a mais do que o sábado; a quarta-feira espera 73.908 ligações a mais; a quinta-feira, 73.082 a mais; e a sexta-feira, 61.898 ligações a mais do que o dia básico.

Com essas informações, é possível elaborar um gráfico – apresentado na figura 28 a seguir – que permita a melhor visualização da sazonalidade do dia da semana.

Figura 28 – Comportamento sazonal do dia da semana – modelo modificado (sem V+3) para prever a quantidade de ligações, produto 103



Fonte: Gráfico elaborado pelo pesquisador, a partir do relatório de regressão do Excel.

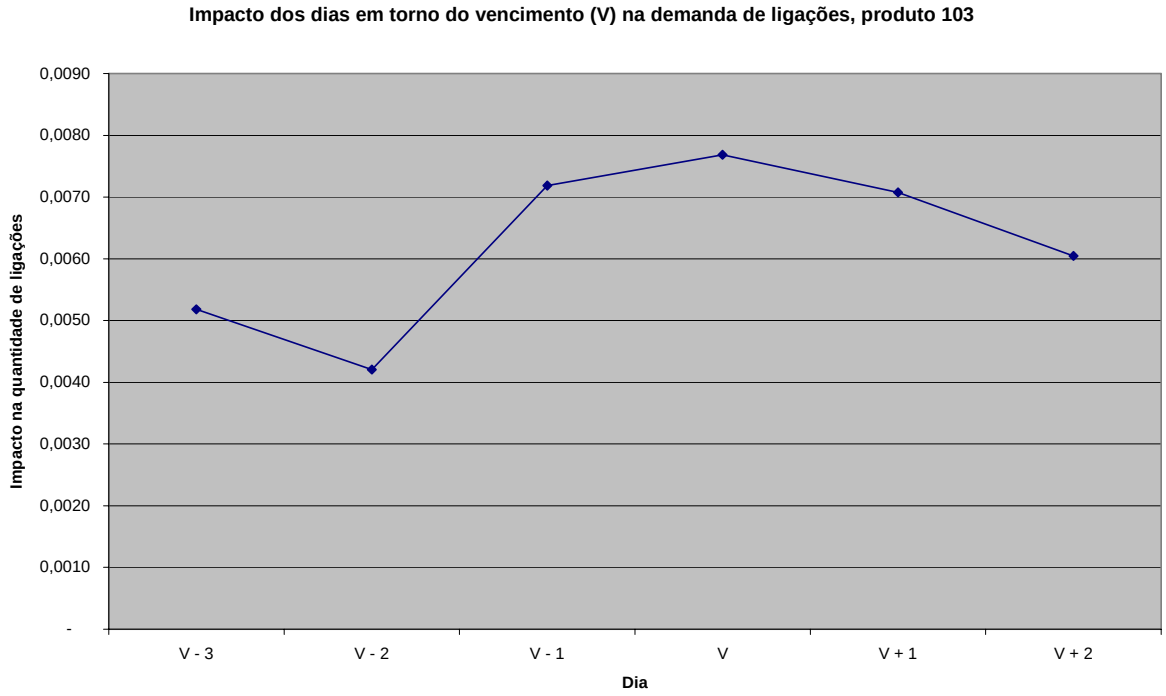
A evolução da demanda de ligações ao longo da semana (considerando dias comuns, ou seja, sem a ocorrência de feriados ou proximidade com chegada ou vencimento de contas) começa com um valor bem baixo no domingo e apresenta um salto bem grande na segunda-feira; a demanda vai caindo suavemente ao longo da semana, tendo uma queda um pouco mais brusca na sexta-feira e bem mais intensa no sábado, antes de cair um pouco menos para voltar ao patamar de domingo.

De forma análoga aos dias da semana, o fato de um dia ser feriado reduz em 71.345 a quantidade prevista de ligações; e cada conta telefônica prevista para chegar à casa do cliente naquele dia aumenta em 0,0034 a previsão da quantidade de ligações demandadas para o dia em questão (assim e se, por exemplo, 100.000 contas estiverem chegando em um dia específico, é de se esperar que isso implique em um aumento de 100.000 x 0,0034 = 340 ligações no contingente diário). Se esse número for examinado de forma percentual, é possível concluir que, em média, 0,34% (= 0,0034) dos clientes ligam para o Telemar no dia em que recebem a sua conta telefônica.

Em relação aos dias que circundam a data do vencimento, é análogo o raciocínio que explica as conseqüências: cada conta vencendo 3 dias depois (ou seja, para as quais a data em questão representa V – 3) aumenta em 0,0052 a quantidade de ligações; cada conta vencendo 2 dias depois aumenta em 0,0042 a quantidade de ligações; cada conta vencendo no dia seguinte aumenta em 0,0072 essa quantidade; cada conta vencendo no dia em questão aumenta em 0,0077 a mesma quantidade; cada conta vencendo na véspera aumenta em 0,0071 a quantidade; e cada conta vencendo na antevéspera aumenta em 0,0060 a quantidade prevista de ligações.

Com essas informações, é possível elaborar um gráfico – apresentado na figura 29 a seguir – que permita a melhor visualização do comportamento da demanda de ligações nos dias em torno da data de vencimento.

Figura 29 – Impacto dos dias em torno do vencimento – modelo modificado (sem V+3) para prever a quantidade de ligações, produto 103



Fonte: Gráfico elaborado pelo pesquisador, a partir do relatório de regressão do Excel.

É possível perceber que o maior impacto ocorre no dia do vencimento. Com exceção de V – 2 (variável cujo coeficiente angular teve valor p acima de 5%), o impacto do dia em questão vai

diminuindo na medida em que a data se afasta do vencimento, para ambos os lados, revelando uma queda mais acentuado para os dias anteriores à data crítica.

Naturalmente, os coeficientes não têm apenas significado isolado, mas também podem e devem ser analisados em conjunto: se as conseqüências de todos os efeitos (dia da semana, ocorrência de feriado, data crítica em relação à chegada ou vencimento da conta) forem incorporadas ao valor previsto para um dia básico, será possível estabelecer uma equação para prever a quantidade demanda de ligações em um determinado dia em função das características da data em questão. Dessa maneira, e nesse caso, a equação pode ser escrita conforme apresentada na figura 30 a seguir.

Figura 30 – Equação do modelo modificado (sem V+3) para prever a quantidade de ligações, produto 103

Quantidade de ligações = 92.693 – 41.103 x **domingo** + 79.874 x **segunda-feira** + 75.900 x **terça-feira** + 73.908 x **quarta-feira** + 73.082 x **quinta-feira** + 61.898 x **sexta-feira** – 71.345 x **feriado** + 0,0034 x **C** + 0,0052 x **V - 3** + 0,0042 x **V - 2** + 0,0072 x **V - 1** + 0,0077 x **V** + 0,0071 x **V + 1** + 0,0060 x **V + 2**

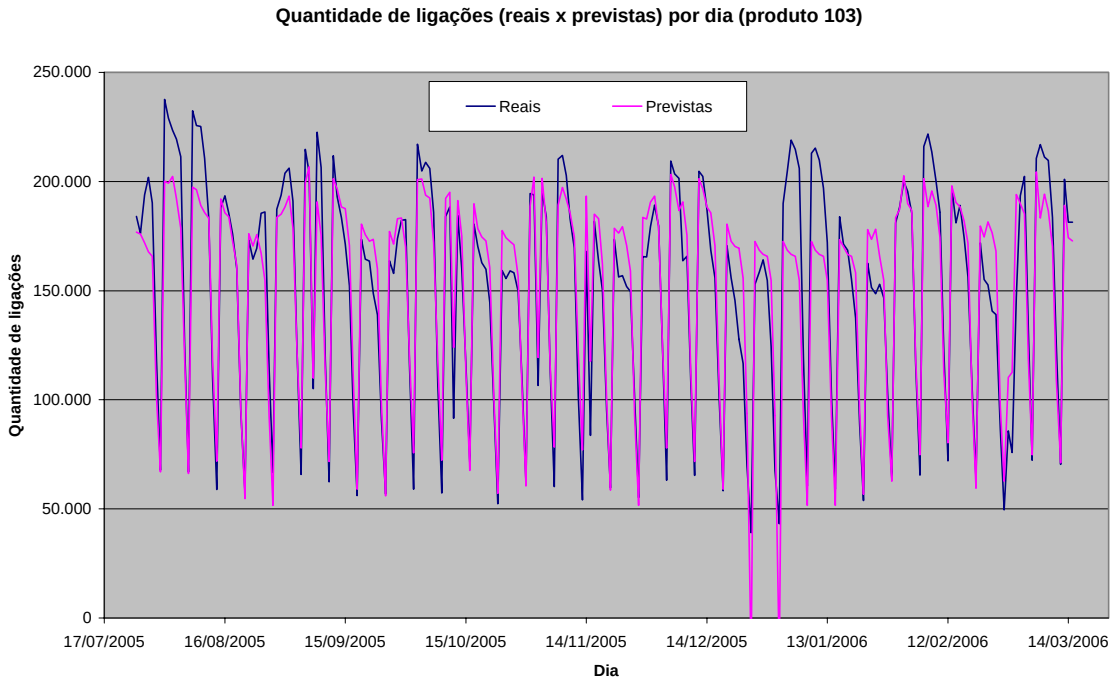
Fonte: Equação elaborada pelo pesquisador, a partir do relatório de regressão do Excel.

onde:

- as 7 primeiras variáveis são do tipo *dummy*, podendo assumir o valor 0 (não) ou 1 (sim);
- as 7 últimas representam, respectivamente: a quantidade de contas chegando no dia; vencendo 3 dias depois; 2 dias depois; 1 dia depois; no próprio dia; 1 dia antes; e 2 dias antes.

De posse dessa equação, então, é possível calcular quanto o modelo de regressão teria previsto para a quantidade de ligações em cada um dos dias acerca dos quais o histórico disponibiliza o valor real desta grandeza. Desta forma, pode-se comparar os valores que teriam sido previstos com os que efetivamente ocorreram, de maneira que um erro de previsão possa ser mensurado. Essa comparação pode ser visualizada na figura 31 a seguir.

Figura 31 – Valores reais (multiplicados por uma constante) e previstos (pelo modelo) da quantidade de ligações por dia, para o produto 103, julho de 2005 a março de 2006



Fonte: Gráfico elaborado pelo pesquisador

A menos de alguns pontos específicos, pode ser percebida uma boa aderência visual do modelo em relação aos dados reais. Mas é necessário formalizar e quantificar essa aderência. A medida de erro escolhida para tal propósito foi o MAPE (erro médio percentual absoluto), anteriormente descrito na seção 3.1.

Para cada dia dentro do período de julho de 2005 a março de 2006, foi organizada uma tabela apresentando a quantidade real de ligações, os valores que o modelo de regressão teria previsto para esta grandeza, assim como os erros percentuais absolutos oriundos desta previsão (em conjunto com a sua média). A tabela 26 a seguir apresenta parte destes dados.

Em média, o modelo está cometendo um erro de previsão da ordem de 10,98%. Esse valor não é muito alto, mas é substancialmente maior que os 5% que a Coordenação de Planejamento costuma cometer na previsão de demanda do produto 103. No entanto, algumas observações são pertinentes, neste momento.

Tabela 26 – Valores reais, previstos e erros do modelo modificado (sem V+3) de regressão múltipla para prever a quantidade de ligações, produto 103, dezembro de 2005

Dia	Dia da semana	Feriado ?	Ligações	Ligações (previstas pelo modelo)	Erro de previsão
01/12/2005	5a. feira	Não	189.179	193.288	2,2%
02/12/2005	6a. feira	Não	179.607	178.716	0,5%
03/12/2005	Sábado	Não	119.738	121.003	1,1%
04/12/2005	Domingo	Não	63.118	77.992	23,6%
05/12/2005	2a. feira	Não	209.398	203.208	3,0%
06/12/2005	3a. feira	Não	203.525	197.677	2,9%
07/12/2005	4a. feira	Não	201.497	186.681	7,4%
08/12/2005	5a. feira	Não	163.731	190.634	16,4%
09/12/2005	6a. feira	Não	165.708	175.358	5,8%
10/12/2005	Sábado	Não	108.033	112.476	4,1%
11/12/2005	Domingo	Não	65.321	71.680	9,7%
12/12/2005	2a. feira	Não	204.689	201.258	1,7%
13/12/2005	3a. feira	Não	202.289	196.754	2,7%
14/12/2005	4a. feira	Não	187.186	188.341	0,6%
15/12/2005	5a. feira	Não	168.579	185.651	10,1%
16/12/2005	6a. feira	Não	155.912	170.652	9,5%
17/12/2005	Sábado	Não	102.895	109.900	6,8%
18/12/2005	Domingo	Não	58.292	59.078	1,3%
19/12/2005	2a. feira	Não	170.636	180.520	5,8%
20/12/2005	3a. feira	Não	155.824	172.655	10,8%
21/12/2005	4a. feira	Não	145.547	170.260	17,0%
22/12/2005	5a. feira	Não	127.727	169.413	32,6%
23/12/2005	6a. feira	Não	116.546	155.684	33,6%
24/12/2005	Sábado	Não	69.839	93.699	34,2%
25/12/2005	Domingo	Sim	39.038	(18.895)	148,4%
26/12/2005	2a. feira	Não	152.909	172.567	12,9%
27/12/2005	3a. feira	Não	158.507	168.593	6,4%
28/12/2005	4a. feira	Não	164.109	166.601	1,5%
29/12/2005	5a. feira	Não	154.690	165.775	7,2%
30/12/2005	6a. feira	Não	124.406	154.591	24,3%
31/12/2005	Sábado	Não	67.541	92.693	37,2%

Fonte: Planilha elaborada pelo pesquisador

É válido lembrar, por exemplo, que, durante o processo-padrão de previsão, os analistas recebem os valores previstos pelo TotalView e acrescentam a eles os impactos esperados por eventos e premissas, como o próprio comportamento do ciclo de contas e a ocorrência de feriados, mas também como outros eventos especiais relacionados ao produto (plano de minutos, campanhas na mídia, crescimento de base e outros potenciais aspectos impactantes,

relacionados no Anexo I). Os 5% de erro de previsão só são conseguidos após a consideração de todas essas subjetividades.

Ora, o modelo de regressão proposto visa substituir apenas os valores sugeridos pelo TotalView, acrescidos da influência dos feriados e do comportamento dos ciclos de contas; a idéia, em nenhuma hipótese, é substituir o papel do analista encarregado da previsão – dotado de toda a sua insubstituível experiência e subjetividade – mas de fornecer-lhe subsídio complementar para a tomada de decisão. Assim, e já que o analista vai poder ser capaz de acrescentar aos valores sugeridos pelo modelo de regressão os impactos que julgar relevantes, é mais justo comparar os erros de previsão do modelo aqui proposto com os erros oriundos do módulo de previsão do TotalView, com o intuito de verificar a ocorrência ou não de melhorias no processo de previsão do produto 103. Tal comparação está apresentada na tabela 27 a seguir.

Com a apresentação dos dados apenas referentes ao período de 1 mês, não é possível visualizar adequadamente os ganhos de acurácia do modelo proposto em relação ao módulo de previsão do TotalView. Entretanto, ao se agrupar os dados referentes a todo o período utilizado para a previsão, é possível calcular que o TotalView cometeu – em média – um erro da ordem de 13,54%, razoavelmente superior aos 10,98% de erro cometidos pelo modelo de regressão, o que confirma a expectativa de ganho de acurácia proporcionado por este último.

Uma análise mais detalhada dos erros de previsão revela que alguns dos dias mais difíceis de prever (maiores erros) consistem em feriados ou "quase-feriados". Por exemplo, os dias 24/12/2005 e 31/12/2005 não foram classificados como feriados, mas, certamente não consistem em dias comuns e devem gerar uma demanda de ligações inferior ao de um dia-padrão. O problema é que o modelo de regressão não levou em conta isso, já que era possível classificar um dia apenas como sendo feriado ou não; e já que esses dias não eram efetivamente feriados, foram tratados pelo modelo como dias comuns. Muito provavelmente por causa disso, suas demandas foram superestimadas e seus erros de previsão foram bem acima da média: 34,2% e 37,2%, respectivamente.

Problemas semelhantes (talvez em menor escala) certamente ocorreram em alguns dias do ano que consistiram em feriados locais, mas não nacionais. Como o modelo propõe uma previsão consolidada (nacional) da demanda, só foi possível considerar os feriados nacionais; mas o dia 20/01/2006 (feriado local na cidade do Rio de Janeiro), por exemplo, teve uma redução na

demanda nacional provocada pela imensa redução na demanda da cidade do Rio de Janeiro e foi tratado como um dia normal pelo modelo, tendo tido a sua demanda superestimada. Esse problema pode ser estendido a dias "enforcados" e "feriadões", em que a demanda foge do padrão.

Tabela 27 – Valores reais, previstos pelo modelo e pelo TotalView e erros de previsão dos mesmos para a quantidade de ligações, produto 103, dezembro de 2005

Dia	Ligações	Ligações (previstas pelo modelo)	**Erro de previsão do modelo**	Ligações (previstas pelo Total View)	**Erro de previsão do Total View**
01/12/2005	189.179	193.288	**2,2%**	192.523	**1,8%**
02/12/2005	179.607	178.716	**0,5%**	178.987	**0,3%**
03/12/2005	119.738	121.003	**1,1%**	109.541	**8,5%**
04/12/2005	63.118	77.992	**23,6%**	61.445	**2,7%**
05/12/2005	209.398	203.208	**3,0%**	195.859	**6,5%**
06/12/2005	203.525	197.677	**2,9%**	191.408	**6,0%**
07/12/2005	201.497	186.681	**7,4%**	191.744	**4,8%**
08/12/2005	163.731	190.634	**16,4%**	185.830	**13,5%**
09/12/2005	165.708	175.358	**5,8%**	172.731	**4,2%**
10/12/2005	108.033	112.476	**4,1%**	105.630	**2,2%**
11/12/2005	65.321	71.680	**9,7%**	59.255	**9,3%**
12/12/2005	204.689	201.258	**1,7%**	189.071	**7,6%**
13/12/2005	202.289	196.754	**2,7%**	184.898	**8,6%**
14/12/2005	187.186	188.341	**0,6%**	185.240	**1,0%**
15/12/2005	168.579	185.651	**10,1%**	176.388	**4,6%**
16/12/2005	155.912	170.652	**9,5%**	154.769	**0,7%**
17/12/2005	102.895	109.900	**6,8%**	115.442	**12,2%**
18/12/2005	58.292	59.078	**1,3%**	81.185	**39,3%**
19/12/2005	170.636	180.520	**5,8%**	165.469	**3,0%**
20/12/2005	155.824	172.655	**10,8%**	153.659	**1,4%**
21/12/2005	145.547	170.260	**17,0%**	126.463	**13,1%**
22/12/2005	127.727	169.413	**32,6%**	127.815	**0,1%**
23/12/2005	116.546	155.684	**33,6%**	150.233	**28,9%**
24/12/2005	69.839	93.699	**34,2%**	113.170	**62,0%**
25/12/2005	39.038	(18.895)	**148,4%**	81.365	**108,4%**
26/12/2005	152.909	172.567	**12,9%**	173.213	**13,3%**
27/12/2005	158.507	168.593	**6,4%**	161.179	**1,7%**
28/12/2005	164.109	166.601	**1,5%**	134.644	**18,0%**
29/12/2005	154.690	165.775	**7,2%**	138.972	**10,2%**
30/12/2005	124.406	154.591	**24,3%**	158.410	**27,3%**
31/12/2005	67.541	92.693	**37,2%**	117.105	**73,4%**

Fonte: Planilha elaborada pelo pesquisador

Um problema ainda mais grave ocorreu com o dia 25/12/2005. Seu erro de previsão foi de quase 150%, já que a sua demanda prevista foi de -18.895 ligações (um valor negativo!). Isso aconteceu porque o dia em questão apresentava dois efeitos negativos (redutores de demanda), que foram "somados" pelo modelo: ser domingo **e** ser feriado. Mas, na prática, o

efeito de um dia ser feriado certamente é reduzido se ele cair em um domingo (ou sábado); e o modelo não levou isso em conta, até porque só é possível considerar um dia como sendo feriado ou não, e não como um "feriado moderado".

Na verdade, esses problemas fizeram com que os erros de previsão aumentassem, mas isso não quer dizer que o processo futuro de previsão terá que incorrer nesses erros de grande magnitude quando ocorrerem dias "problemáticos". Quando os analistas de previsão estiverem utilizando o modelo aqui proposto, eles poderão "manipular" os valores da variável *dummy* **feriado** para os dias problemáticos.

Assim, por exemplo, o dia 24/12 pode receber manualmente o valor 0,7 para a sua variável feriado, ao invés de ser tratado como um dia normal; o dia 20/01 pode receber o valor 0,4 da mesma forma; um feriado que caia no domingo pode ter o seu valor da variável feriado reduzido de 1 para 0,6.

Depois desta manipulação, basta considerar os *inputs* do setor de faturamento da Telemar (informações sobre os ciclos de contas) e usar a equação apresentada anteriormente na figura 30 para prever a quantidade de ligações para o produto 103, a cada dia.

Em cima do *output* da equação, então, os analistas devem incorporar os impactos esperados pelos eventos especiais (campanhas de mídias, novos serviços etc.) e, dessa forma, os erros de previsão tenderão a diminuir substancialmente.

6.1.3. O modelo de previsão para o tempo médio de atendimento (TMA)

Além da demanda de ligações, também foi necessário tentar explicar o comportamento do tempo médio de atendimento (TMA) do produto 103, já que esta variável é um *input* fundamental para a atividade de dimensionamento da capacidade e a sua previsão consiste em uma tarefa muito importante da Coordenação de Planejamento de Tráfego.

Assim, de uma forma análoga, foi montada uma planilha de Excel com os dados necessários para a construção do modelo de regressão múltipla dentro do qual o comportamento da variável dependente "tempo médio de atendimento em um dia específico" poderia estar sendo explicado pelas mesmas 9 variáveis independentes do modelo modificado (já desconsiderando V + 3) anterior: dia da semana; feriado; chegada (C); vencimento (V); V - 3; V - 2; V - 1; V + 1; e V + 2.

A tabela 28 a seguir apresenta parcialmente a organização da planilha de dados em questão.

Tabela 28 – Planilha de dados do modelo de regressão múltipla para prever o TMA, produto 103, agosto de 2005

Dia	Dia da semana	Feriado ?	Chegada (C)	V - 3	V - 2	V - 1	Vencimento (V)	V + 1	V + 2	TMA (seg)
01/08/2005	2a. feira	Não	-	2.142.215	-	-	2.136.197	-	-	145,50
02/08/2005	3a. feira	Não	1.918.920	-	2.142.215	-	-	2.136.197	-	144,00
03/08/2005	4a. feira	Não	2.177.552	-	-	2.142.215	-	-	2.136.197	144,00
04/08/2005	5a. feira	Não	-	1.918.920	-	-	2.142.215	-	-	147,00
05/08/2005	6a. feira	Não	218.438	-	1.918.920	-	-	2.142.215	-	144,75
06/08/2005	Sábado	Não	-	-	-	1.918.920	-	-	2.142.215	129,00
07/08/2005	Domingo	Não	-	-	-	-	1.918.920	-	-	113,25
08/08/2005	2a. feira	Não	-	2.177.552	-	-	-	1.918.920	-	146,25
09/08/2005	3a. feira	Não	2.036.848	-	2.177.552	-	-	-	1.918.920	148,50
10/08/2005	4a. feira	Não	801.143	798.630	-	2.177.552	-	-	-	149,25
11/08/2005	5a. feira	Não	-	-	798.630	-	2.177.552	-	-	147,00
12/08/2005	6a. feira	Não	-	1.456.656	-	798.630	-	2.177.552	-	141,75
13/08/2005	Sábado	Não	-	-	1.456.656	-	798.630	-	2.177.552	125,25
14/08/2005	Domingo	Não	-	801.143	-	1.456.656	-	798.630	-	110,25
15/08/2005	2a. feira	Não	-	-	801.143	-	1.456.656	-	798.630	138,75
16/08/2005	3a. feira	Não	296.642	-	-	801.143	-	1.456.656	-	142,50
17/08/2005	4a. feira	Não	246.901	296.642	-	-	801.143	-	1.456.656	143,25
18/08/2005	5a. feira	Não	-	-	296.642	-	-	801.143	-	142,50
19/08/2005	6a. feira	Não	-	-	-	296.642	-	-	801.143	142,50
20/08/2005	Sábado	Não	-	246.901	-	-	296.642	-	-	123,75
21/08/2005	Domingo	Não	-	-	246.901	-	-	296.642	-	109,50
22/08/2005	2a. feira	Não	-	-	-	246.901	-	-	296.642	139,50
23/08/2005	3a. feira	Não	-	-	-	-	246.901	-	-	139,50
24/08/2005	4a. feira	Não	2.136.197	-	-	-	-	246.901	-	139,50
25/08/2005	5a. feira	Não	-	-	-	-	-	-	246.901	139,50
26/08/2005	6a. feira	Não	-	-	-	-	-	-	-	133,50
27/08/2005	Sábado	Não	-	-	-	-	-	-	-	118,50
28/08/2005	Domingo	Não	-	-	-	-	-	-	-	100,50
29/08/2005	2a. feira	Não	-	2.136.197	-	-	-	-	-	139,50
30/08/2005	3a. feira	Não	2.181.734	-	2.136.197	-	-	-	-	139,50
31/08/2005	4a. feira	Não	1.879.401	-	-	2.136.197	-	-	-	145,50

Fonte: Planilha elaborada pelo pesquisador

Também neste caso, as variáveis independentes “dia da semana” e “feriado” foram trabalhadas como sendo do tipo *dummy*; e, novamente através da ferramenta de “Regressão” do suplemento de “Análise de Dados” do Excel, foi gerado o modelo de regressão múltipla proposto. O relatório está apresentado nas tabelas 29, 30 e 31 a seguir.

Tabela 29 – Estatística de regressão do modelo de regressão múltipla para prever o TMA, produto 103

Estatística de regressão	
R múltiplo	0,89
R-Quadrado	0,80
R-quadrado ajustado	**0,79**
Erro padrão	6,7110
Observações	234

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

A primeira análise aponta para um modelo com um potencial muito bom, já que 79% (R-quadrado ajustado) da variação dos dados pode ser explicada por ele. Para analisar o F de significação, é preciso examiná-lo a partir da tabela 30 a seguir.

Tabela 30 – Análise de Variância do modelo de regressão múltipla para prever a o TMA, produto 103

ANOVA

	gl	*F*	*F de significação*
Regressão	14,00	62,76	**0,0000**
Resíduo	219,00		
Total	233,00		

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

O F de significação resultante foi 0,0000 (zero). Assim, tende-se a acreditar na real influência das variáveis explicativas no tempo médio de atendimento e aceitar o modelo com bastante certeza.

Para analisar a contribuição de cada variável independente, é necessário examinar seus coeficientes, juntamente com os respectivos valores p, apresentados na tabela 31 a seguir.

Ao analisá-la, é possível perceber um altíssimo valor p para a variável Chegada (C), indicando que a "quantidade de contas chegando ao cliente naquele dia específico" não está funcionando como um bom explicador linear para o TMA.

Apesar de ser uma variável crítica, ela deve ser retirada da análise, pois seu valor p está muito acima do razoável (10%) e porque o sinal do seu coeficiente angular (negativo) estaria indicando que o tempo médio de atendimento em um determinado dia diminuiria para cada conta que chegasse ao cliente no dia em questão, constituindo um efeito contrário ao esperado.

As variáveis V – 3 e a V – 2 também apresentam valores p altos, indicando que a "quantidade de contas para as quais aquele dia específico corresponde a 2 ou 3 dias antes do vencimento" não está funcionando como um bom explicador linear para o TMA. O efeito "pré-vencimento" para o TMA parece, então, começar apenas na véspera (V – 1) .

Tabela 31 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo de regressão múltipla para prever o TMA, produto 103

	Coeficientes	*valor-P*	*95% inferiores*	*95% superiores*
Interseção	**122,27**	0,00%	120	125
Domingo	**(15,17)**	**0,00%**	(18)	(12)
2a. Feira	**15,37**	**0,00%**	12	19
3a. Feira	**15,20**	**0,00%**	12	19
4a. Feira	**17,70**	**0,00%**	14	21
5a. Feira	**17,86**	**0,00%**	15	21
6a. Feira	**15,11**	**0,00%**	12	18
Feriado	**(29,20)**	**0,00%**	(34)	(24)
Chegada (C)	**(0,0000003)**	**68,35%**	(0,0000)	0,0000
V - 3	**0,0000008**	**29,97%**	(0,0000)	0,0000
V - 2	**0,0000011**	**16,89%**	(0,0000)	0,0000
V - 1	**0,0000016**	**3,44%**	0,0000	0,0000
Vencimento (V)	**0,0000029**	**0,01%**	0,0000	0,0000
V + 1	**0,0000026**	**0,12%**	0,0000	0,0000
V + 2	**0,0000022**	**0,36%**	0,0000	0,0000

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

Assim como a variável Chegada (C), as variáveis V – 3 e V – 2 podem estar "poluindo" o modelo e convém retirá-las da análise. A ferramenta de "Regressão" do Excel foi novamente usada para gerar o modelo de regressão múltipla. As tabelas 32, 33 e 34 a seguir foram extraídas do relatório fornecido pela ferramenta.

Tabela 32 – Estatística de regressão do modelo modificado (sem C; V – 3; V – 2) de Regressão múltipla para prever o TMA, produto 103

Estatística de regressão	
R múltiplo	0,89
R-Quadrado	0,80
R-quadrado ajustado	**0,79**
Erro padrão	6,7012
Observações	234

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

O modelo continuou com o mesmo potencial: 79% da variação dos dados pode ser explicada por ele. O F de significação pode ser observado na tabela 33 a seguir.

Tabela 33 – Análise de Variância do modelo modificado (sem C; V – 3; V – 2) de regressão múltipla para prever o TMA, produto 103

ANOVA

	gl	*F*	*F de significação*
Regressão	11,00	79,89	**0,0000**
Resíduo	222,00		
Total	233,00		

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

O F de significação resultante novamente foi 0,0000 (zero). Também neste caso, pode-se acreditar na real influência das variáveis explicativas no tempo médio de atendimento e, portanto, aceitar o modelo com bastante certeza.

Os coeficientes linear e angulares, juntamente com os respectivos valores p, estão apresentados na tabela 34 a seguir.

Tabela 34 – Valor, representatividade e intervalo de confiança dos coeficientes das variáveis independentes do modelo modificado (sem C; V – 3; V – 2) de regressão múltipla para prever o TMA, produto 103

	Coeficientes	*valor-P*	*95% inferiores*	*95% superiores*
Interseção	**122,64**	0,00%	120	125
Domingo	**(15,15)**	**0,00%**	(18)	(12)
2a. Feira	**15,36**	**0,00%**	12	19
3a. Feira	**14,96**	**0,00%**	12	18
4a. Feira	**17,33**	**0,00%**	14	21
5a. Feira	**17,84**	**0,00%**	15	21
6a. Feira	**15,03**	**0,00%**	12	18
Feriado	**(28,89)**	**0,00%**	(34)	(24)
V - 1	**0,0000012**	**6,96%**	(0,0000)	0,0000
Vencimento (V)	**0,0000031**	**0,00%**	0,0000	0,0000
V + 1	**0,0000032**	**0,00%**	0,0000	0,0000
V + 2	**0,0000024**	**0,04%**	0,0000	0,0000

Fonte: Relatório de regressão do Excel, elaborado pelo pesquisador

Como pode ser observado, os valores p das outras variáveis não mudaram muito e, para este modelo modificado, apenas uma variável apresenta coeficiente angular com valor p superior (mas não muito) a 5%: V – 1; ela, no entanto, será mantida no modelo, porque o valor p do seu coeficiente angular está abaixo de 10%.

Desta forma, a retirada das variáveis C, V – 3 e V – 2 foi benéfica para o modelo (tornando-o, inclusive, mais enxuto) e os seus coeficientes podem ser, então, interpretados.

O coeficiente linear da regressão apresentou o valor 122,64. Dessa forma – e de acordo com o que foi explicado anteriormente, nesta seção, para o modelo da quantidade de ligações – quando o dia for sábado, não-feriado e não consistir de nenhuma data dentro do intervalo de 4 dias em torno do vencimento, a previsão é que o tempo médio de atendimento (TMA) seja 122,64 segundos.

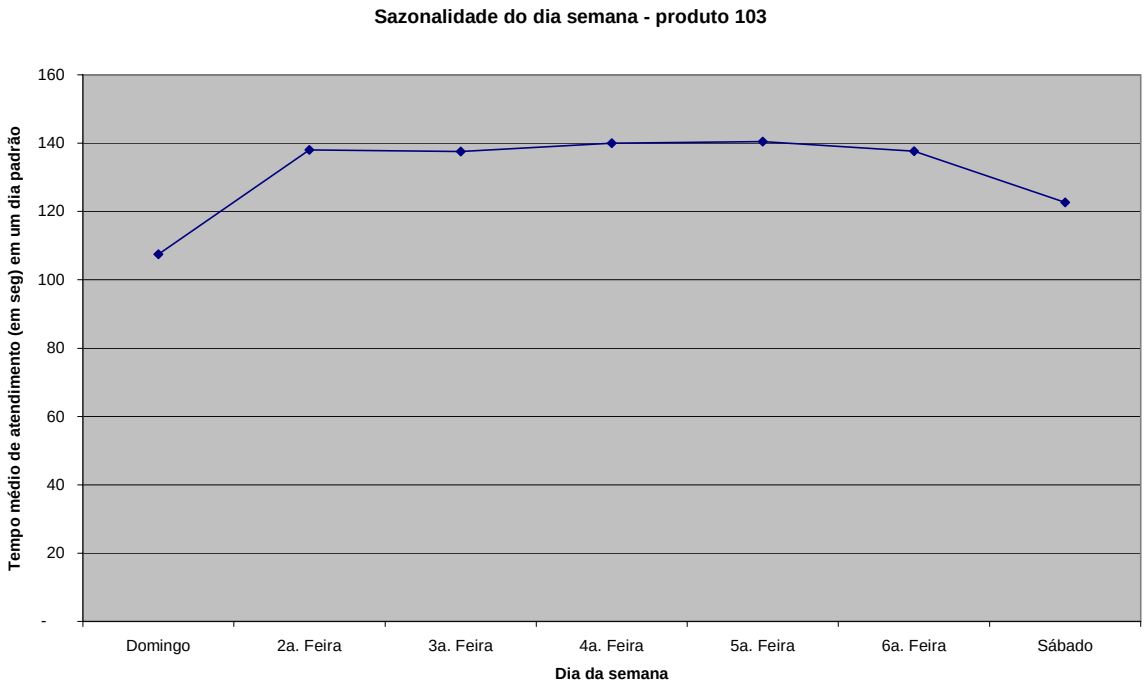
Esse número pode ser encarado como uma quantidade básica, à qual devem ser adicionados os efeitos da ocorrência das variáveis independentes, expressados através dos seus respectivos coeficientes angulares, destacados na tabela 34 anterior.

Assim, o fato do dia ser domingo contribui com uma diminuição de 15,15 segundos no TMA previsto em relação ao número básico (122,64); o fato de o dia ser segunda-feira, com um

aumento esperado de 15,36 segundos no TMA; do dia ser terça-feira, com um aumento de 14,96 segundos; de ser quarta-feira, com um aumento de 17,33 segundos; de ser quinta-feira, com 17,84 segundos a mais; e de ser sexta-feira, contribui com um aumento de 15,03 segundos na previsão do tempo médio de atendimento, em relação ao dia básico.

Com essas informações, é possível elaborar um gráfico – apresentado na figura 32 a seguir – que permita a melhor visualização da sazonalidade do dia da semana.

Figura 32 – Comportamento sazonal do dia da semana – modelo modificado (sem C; V – 3; V – 2) para prever o TMA, produto 103



Fonte: Gráfico elaborado pelo pesquisador, a partir do relatório de regressão do Excel.

A evolução da demanda de ligações ao longo da semana (considerando dias comuns) começa com um valor bem baixo no domingo e apresenta um salto bem grande na segunda-feira, revelando, a partir daí, uma estabilidade muito grande ao longo dos outros dias úteis, tendo uma queda relativamente intensa no sábado, antes de cair um pouco menos para voltar ao patamar de domingo.

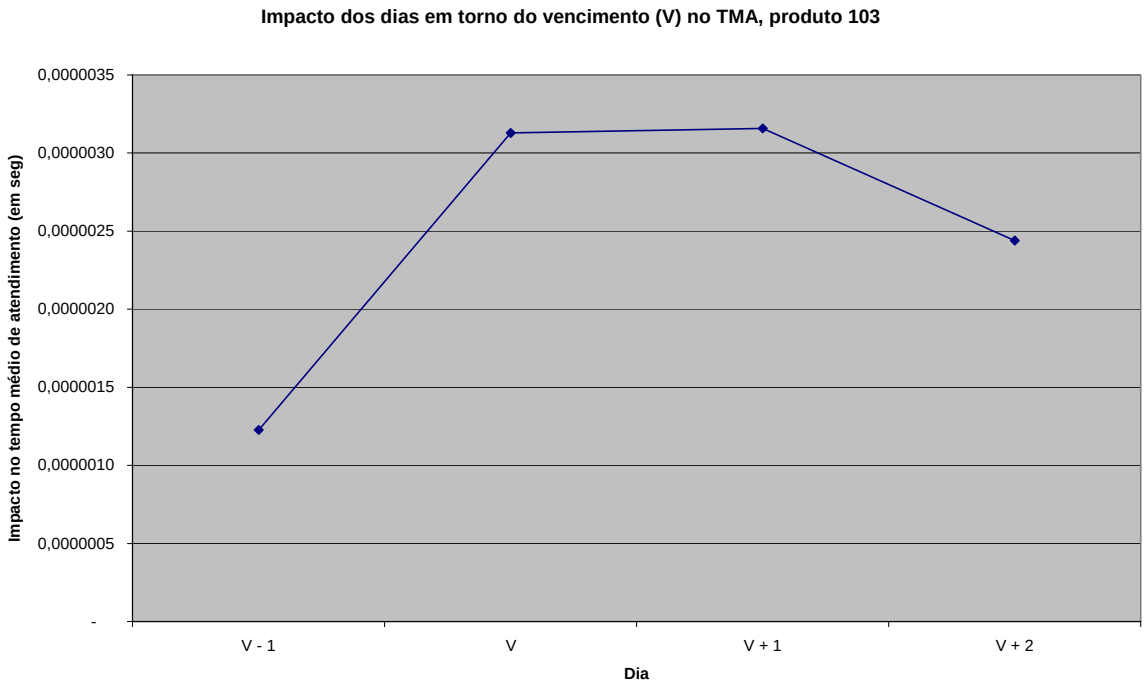
De forma análoga aos dias da semana, o fato de um dia ser feriado reduz em 28,89 segundos o tempo médio de atendimento previsto. Essa grande redução (de até 30 segundos) no tempo médio de atendimento tanto para feriados como para finais de semana quando comparados

aos dias úteis é um tanto quanto estranha. Em princípio, não foi possível especular os motivos que levariam a tal variação. A equipe de previsão foi indagada a respeito e também não foi capaz de justificar o fato. A investigação acerca de tais motivos pode constituir-se em um interessante objeto de pesquisa futura.

Em relação aos dias em torno da data de vencimento, os seguintes efeitos podem ser explicados: cada conta vencendo no dia seguinte (ou seja, para as quais a data em questão representa V – 1) aumenta em 0,0000012 segundos o tempo médio de atendimento; cada conta vencendo no dia em questão aumenta em 0,0000031 segundos o TMA; cada conta vencendo na véspera aumenta em 0,0000032 segundos essa grandeza; e cada conta vencendo na antevéspera aumenta em 0,0000024 segundos o tempo médio de atendimento.

Com essas últimas informações, é possível elaborar um gráfico – apresentado na figura 33 a seguir – que permita a melhor visualização do comportamento do TMA nos dias em torno da data de vencimento.

Figura 33 – Impacto dos dias em torno do vencimento – modelo modificado (sem C; V – 3; V – 2) para prever o TMA, produto 103



Fonte: Gráfico elaborado pelo pesquisador, a partir do relatório de regressão do Excel.

É possível perceber que o maior impacto ocorre no dia do vencimento e no dia seguinte; e que a véspera do vencimento impacta substancialmente menos (até em comparação ao V + 2) o tempo médio de atendimento.

Se as conseqüências de todos os efeitos (dia da semana, ocorrência de feriado, data crítica em relação ao vencimento da conta) forem incorporadas ao valor previsto para um dia básico, será possível estabelecer uma equação para prever o tempo médio de atendimento (TMA) em um determinado dia em função das características da data em questão. Dessa maneira, e nesse caso, a equação pode ser escrita conforme consta na figura 34 a seguir.

Figura 34 – Equação do modelo modificado (sem C; V – 3; V – 2) para prever o TMA, produto 103

TMA (em segundos) = 122,64 – 15,15 x **domingo** + 15,36 x **segunda-feira** + 14,96 x **terça-feira** + 17,33 x **quarta-feira** + 17,84 x **quinta-feira** + 15,03 x **sexta-feira** – 28,89 x **feriado** + 0,0000012 x **V - 1** + 0,0000031 x **V** + 0,0000032 x **V + 1** + 0,0000024 x **V + 2**

Fonte: Equação elaborada pelo pesquisador, a partir do relatório de regressão do Excel.

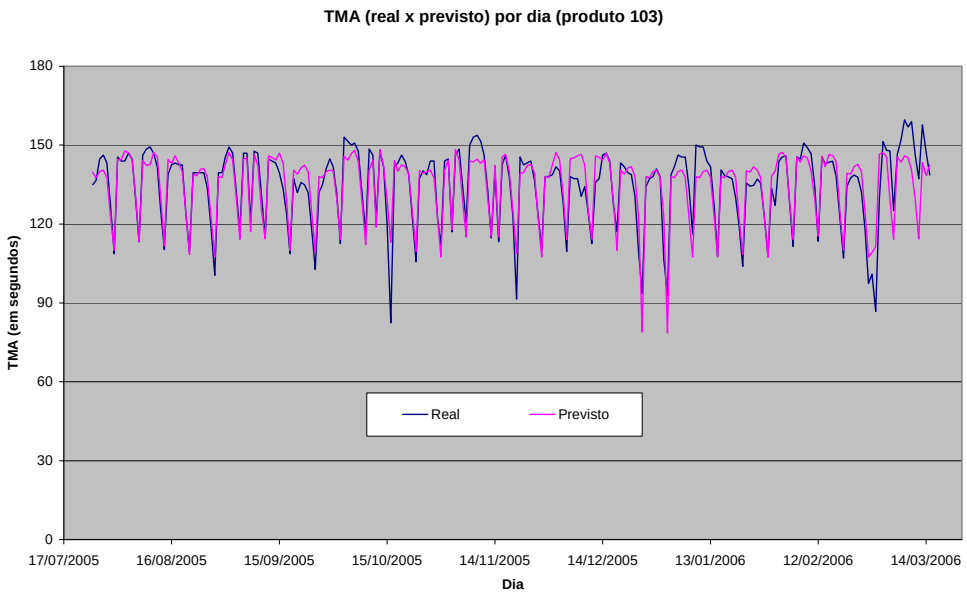
onde:

- as 7 primeiras variáveis são do tipo *dummy*, podendo assumir o valor 0 (não) ou 1 (sim);
- as 4 últimas representam, respectivamente: a quantidade de contas vencendo 1 dia depois; no próprio dia; 1 dia antes; e 2 dias antes.

Com a equação gerada, pode-se calcular quanto o modelo de regressão teria previsto para o tempo médio de atendimento em cada um dos dias em questão. Assim, é possível comparar os valores que teriam sido previstos com os que efetivamente ocorreram, de forma que um erro de previsão possa ser obtido. Essa comparação pode ser visualizada na figura 35 a seguir.

Pode ser percebida uma aderência visual muito boa do modelo em relação aos dados reais. A medida de erro escolhida também nesse caso para quantificar a aderência foi o MAPE (erro médio percentual absoluto).

Figura 35 – Valores reais (multiplicados por uma constante) e previstos do TMA por dia, para o produto 103, julho de 2005 a março de 2006



Fonte: Gráfico elaborado pelo pesquisador

Para cada dia dentro do período de julho de 2005 a março de 2006, foi organizada uma tabela apresentando o tempo médio de atendimento real, os valores que o modelo de regressão teria previsto para esta grandeza, assim como os erros percentuais absolutos incorridos esta previsão, juntamente com a sua média. A tabela 35 a seguir apresenta parcialmente estes dados.

Em média, o modelo está cometendo um erro de previsão da ordem de 3,61%. Esse valor é satisfatoriamente baixo e bem inferior ao erro médio cometido pelo modelo para prever a quantidade de ligações (quase 11%). Isso revela que o tempo médio de atendimento é uma grandeza mais estável e, portanto, de maior previsibilidade, quando comparada à demanda de ligações.

Tabela 35 – Valores reais, previstos e erros do modelo (sem C; V – 3; V – 2) de regressão múltipla para prever o TMA, produto 103, janeiro de 2006

Dia	Dia da semana	Feriado ?	TMA (seg)	TMA (seg) (previsto pelo modelo)	Erro de previsão
01/01/2006	Domingo	Sim	93,00	78,60	15,5%
02/01/2006	2a. feira	Não	138,75	137,99	0,5%
03/01/2006	3a. feira	Não	141,75	137,60	2,9%
04/01/2006	4a. feira	Não	146,25	139,97	4,3%
05/01/2006	5a. feira	Não	145,50	140,48	3,5%
06/01/2006	6a. feira	Não	145,50	137,67	5,4%
07/01/2006	Sábado	Não	134,25	122,64	8,6%
08/01/2006	Domingo	Não	116,25	107,49	7,5%
09/01/2006	2a. feira	Não	150,00	137,99	8,0%
10/01/2006	3a. feira	Não	149,25	137,60	7,8%
11/01/2006	4a. feira	Não	149,25	139,97	6,2%
12/01/2006	5a. feira	Não	144,00	140,48	2,4%
13/01/2006	6a. feira	Não	141,75	137,67	2,9%
14/01/2006	Sábado	Não	126,75	122,64	3,2%
15/01/2006	Domingo	Não	108,00	107,49	0,5%
16/01/2006	2a. feira	Não	140,53	137,99	1,8%
17/01/2006	3a. feira	Não	138,33	137,60	0,5%
18/01/2006	4a. feira	Não	137,94	139,97	1,5%
19/01/2006	5a. feira	Não	137,13	140,48	2,4%
20/01/2006	6a. feira	Não	129,92	137,67	6,0%
21/01/2006	Sábado	Não	119,52	122,64	2,6%
22/01/2006	Domingo	Não	103,92	108,36	4,3%
23/01/2006	2a. feira	Não	135,62	140,21	3,4%
24/01/2006	3a. feira	Não	134,42	139,84	4,0%
25/01/2006	4a. feira	Não	134,69	141,70	5,2%
26/01/2006	5a. feira	Não	137,07	140,48	2,5%
27/01/2006	6a. feira	Não	135,58	137,67	1,5%
28/01/2006	Sábado	Não	122,99	122,64	0,3%
29/01/2006	Domingo	Não	107,68	107,49	0,2%
30/01/2006	2a. feira	Não	133,53	137,99	3,3%
31/01/2006	3a. feira	Não	127,10	140,22	10,3%

Fonte: Planilha elaborada pelo pesquisador

Da mesma forma (e pelos mesmos motivos) que no caso da previsão da quantidade de ligações, é mais justo comparar os erros de previsão do modelo aqui proposto com os erros oriundos do módulo de previsão do TotalView, com o intuito de verificar a ocorrência ou não de melhorias no processo de previsão do produto 103. Tal comparação está apresentada na tabela 36 a seguir.

Tabela 36 – Valores reais, previstos pelo modelo e pelo TotalView e erros de previsão dos mesmos para o TMA, produto 103, janeiro de 2006

Dia	TMA (seg)	TMA (seg) (previsto pelo modelo)	**Erro de previsão do modelo**	TMA (seg) (previsto pelo Total View)	**Erro de previsão do Total View**
01/01/2006	93,00	78,60	**15,5%**	126,82	**36,4%**
02/01/2006	138,75	137,99	**0,5%**	144,47	**4,1%**
03/01/2006	141,75	137,60	**2,9%**	143,83	**1,5%**
04/01/2006	146,25	139,97	**4,3%**	141,45	**3,3%**
05/01/2006	145,50	140,48	**3,5%**	143,99	**1,0%**
06/01/2006	145,50	137,67	**5,4%**	142,50	**2,1%**
07/01/2006	134,25	122,64	**8,6%**	131,51	**2,0%**
08/01/2006	116,25	107,49	**7,5%**	122,64	**5,5%**
09/01/2006	150,00	137,99	**8,0%**	140,17	**6,6%**
10/01/2006	149,25	137,60	**7,8%**	139,59	**6,5%**
11/01/2006	149,25	139,97	**6,2%**	137,64	**7,8%**
12/01/2006	144,00	140,48	**2,4%**	140,02	**2,8%**
13/01/2006	141,75	137,67	**2,9%**	139,11	**1,9%**
14/01/2006	126,75	122,64	**3,2%**	128,18	**1,1%**
15/01/2006	108,00	107,49	**0,5%**	120,09	**11,2%**
16/01/2006	140,53	137,99	**1,8%**	135,02	**3,9%**
17/01/2006	138,33	137,60	**0,5%**	134,88	**2,5%**
18/01/2006	137,94	139,97	**1,5%**	135,61	**1,7%**
19/01/2006	137,13	140,48	**2,4%**	137,13	**0,0%**
20/01/2006	129,92	137,67	**6,0%**	134,07	**3,2%**
21/01/2006	119,52	122,64	**2,6%**	119,62	**0,1%**
22/01/2006	103,92	108,36	**4,3%**	103,45	**0,5%**
23/01/2006	135,62	140,21	**3,4%**	134,50	**0,8%**
24/01/2006	134,42	139,84	**4,0%**	134,34	**0,1%**
25/01/2006	134,69	141,70	**5,2%**	136,60	**1,4%**
26/01/2006	137,07	140,48	**2,5%**	138,24	**0,9%**
27/01/2006	135,58	137,67	**1,5%**	135,01	**0,4%**
28/01/2006	122,99	122,64	**0,3%**	120,58	**2,0%**
29/01/2006	107,68	107,49	**0,2%**	104,62	**2,8%**
30/01/2006	133,53	137,99	**3,3%**	135,65	**1,6%**
31/01/2006	127,10	140,22	**10,3%**	135,65	**6,7%**

Fonte: Planilha elaborada pelo pesquisador

Com o mesmo formato de apresentação dos dados da comparação entre os erros de previsão para a quantidade de ligações, também não é possível visualizar adequadamente os ganhos de acurácia do modelo proposto em relação ao módulo de previsão do TotalView, para o caso do TMA. Mas, e da mesma forma que anteriormente, é possível calcular que o TotalView cometeu – em média – um erro da ordem de 5,58%, razoavelmente superiores aos 3,61% de erro cometidos pelo modelo de regressão, o que confirma a expectativa de ganho de acurácia proporcionado por este último, também para o tempo médio de atendimento.

As mesmas observações a respeito da ocorrência de "quase-feriados" e de feriados "moderados", realizadas quando da análise dos erros de previsão da demanda de ligações, são também, agora, pertinentes.

Assim como no modelo anterior para prever a quantidade de chamadas para o produto 103, os analistas de previsão:

1. poderão "manipular" os valores da variável *dummy* **feriado** para os dias problemáticos;

2. deverão considerar as informações sobre os ciclos de contas e usar a equação apresentada anteriormente na figura 34 para prever o TMA a cada dia;

3. e poderão, finalmente, incorporar os impactos esperados pelos eventos especiais (campanhas de mídias, novos serviços etc.).

6.2. Dimensionamento da capacidade de atendimento

Conforme descrito na seção 5.2.2, as planilhas Excel usadas para o dimensionamento e escalonamento dos atendentes para o produto 102 utilizam fórmulas Erlang para calcular o nível de serviço (percentual de ligações atendidas em até 10 segundos) em um determinado horário em função do número de operadores dimensionados para o mesmo (levando em conta também a previsão de demanda para tal horário).

A própria equipe de planejamento, no entanto, questiona essa metodologia de cálculo (através de fórmulas Erlang). Para a realidade dos *call centers*, a teoria de filas pode ser a melhor metodologia analítica, mas existem métodos experimentais – como a simulação – que podem ser mais adequados para uma indústria com o cotidiano operacional tão complexo.

A ferramenta de simulação era desconhecida para a equipe que – quando foi brevemente apresentada a esta abordagem alternativa para lidar com o problema de dimensionamento – ficou com a impressão de se tratar de algo muito complexo (nem tanto em termos de saber como utilizá-la, mas mais em termos do que ela realmente faz e como ela lida com o relacionamento entre as entidades envolvidas). A equipe ficou com a impressão que a ferramenta exige *inputs* com um nível de detalhes muito difícil de ser conseguido.

Em vista disso, os responsáveis pelo dimensionamento não se mostraram tão receptivos à possibilidade de utilizar esta nova abordagem para lidar com o problema e nem tão dispostos a "comprar" esta idéia. Sua principal motivação provavelmente não consistiu no medo de uma

possível desvalorização do seu trabalho, mas sim na incapacidade de entender completamente o funcionamento da ferramenta alternativa e de enxergar as seus vantagens.

Caso esse ceticismo acerca dos benefícios potenciais da abordagem experimental venha a ser contornado, a empresa admite a possibilidade de adquirir o *software* necessário para a implementação da mesma, mesmo estando ciente do seu custo elevado. O fato deste ceticismo ainda existir, no entanto, certamente contribuiu para a ocorrência de uma morosidade e de uma certa resistência no envio das informações necessárias para o desenvolvimento da análise do problema de dimensionamento.

Com alguma dificuldade, as informações mais relevantes foram, entretanto, obtidas e puderam ser trabalhadas. O único problema consistiu no fato de não ter sido possível resgatar as informações referentes a cada ligação individual, conforme mencionado na seção 4.4. Mas isso não foi motivado por nenhum outro fator, que não a ausência – por parte dos responsáveis pelo dimensionamento – do conhecimento técnico e do privilégio necessários para acessar a base de dados da Contax.

Não obstante esse ceticismo, a pesquisa empreendida pelo autor e apresentada ao longo das seções 3.3. 3.4 e 3.5 corrobora o argumento da maior adequação dos métodos experimentais a sistemas complexos, como o de um *call center* moderno.

Em especial, e conforme descrito na seção 5.2.2, a metodologia utilizada (Erlang) para o cálculo do nível de serviço não considera a hipótese de abandono por parte do cliente ou sinal de ocupado; ou seja, os clientes que chegam e não encontram agentes disponíveis, aguardam indefinidamente em uma fila até serem atendidos.

Isso cria uma tendência de subestimar o nível de serviço (e, conseqüentemente, superestimar o número de agentes necessários para atingir a meta) porque as fórmulas consideram que o cliente vai esperar indefinidamente. Mas, na operação real, alguns clientes vão abandonar as ligações (se esperarem muito), diminuindo a fila e fazendo com que sejam atendidas em menos tempo as chamadas programadas para serem respondidas após aquela que abandonou a fila, aumentando o nível de serviço.

Adicionalmente, as planilhas usadas para o cálculo do nível de serviço consideram que o tempo de atendimento segue uma distribuição exponencial. No entanto, isso raramente ocorre na prática, caracterizando uma premissa relativamente forçada.

O emprego da simulação permite contemplar as características destacadas nas seções 3.4 e 3.5, incluindo o comportamento de abandono (podendo também considerar que um percentual dos clientes que abandonam irão retornar a ligação, tentando um novo contato dentro de uma quantidade de tempo, que pode ser modelada por uma distribuição estatística) e uma flexibilidade na definição do formato da distribuição do tempo de atendimento.

Para ratificar essa melhor adequação, a idéia consiste em simular computacionalmente – em poucos segundos – a operação do *call center* durante períodos de 30 minutos, contemplando de forma mais realista as suas características para obter resultados mais acurados do que com as metodologias analíticas.

Dessa forma, não é necessário experimentar na prática algumas alternativas de dimensionamento para saber mais acuradamente as suas conseqüências; a experimentação é realizada no ambiente computacional. Mas é possível visualizar a operação em si (com as chamadas chegando, sendo enfileiradas e depois atendidas) e o que estaria acontecendo, de forma detalhada (praticamente como se fosse *in loco*), para poder entender porque determinado período do dia apresentou um nível de serviço tão baixo, por exemplo (ao invés de apenas aceitar o número frio fornecido pelas fórmulas analíticas).

Com o uso da simulação, é possível calcular empiricamente (ao invés de estimar analiticamente) – através do uso de algumas premissas históricas (de previsão de demanda e comportamento dos clientes e do sistema) e da quantidade dimensionada de atendentes como *inputs* – alguns indicadores de performance importantes, como: nível de serviço, taxa de abandono, tempo médio de espera, utilização e ociosidade dos agentes.

Para o dimensionamento e escalonamento dos operadores para atender os clientes *plus* do produto 102 durante o mês de agosto de 2006, foi usada a premissa (oriunda da previsão de demanda) de que chegariam à central 586 chamadas, com um tempo médio de atendimento de 29 segundos, na primeira meia-hora do dia (00:00 às 00:30), conforme pode ser observado na tabela 10 anterior.

De acordo com a mesma tabela, seria necessário (segundo as fórmulas analíticas), para lidar com esse intervalo de 30 minutos, alocar 14 atendentes (já considerando uma redução esperada, por conta do absenteísmo), de forma a atingir 85% de nível de serviço (clientes sendo atendidos em até 10 segundos).

A equipe de dimensionamento alocou 13 atendentes para este horário e as mesmas fórmulas analíticas previram 88,04% de nível de serviço para este período, já levando em conta a premissa de que 4% dos operadores costumam faltar em dias comuns; ou seja, dimensionar 13 atendentes consiste em contar com 13 * (1 – 4%) = 12 atendentes, aproximadamente.

Com o objetivo de questionar esses números, foi construído – no *software* Arena Contact Center – um modelo para simular como se comportaria o sistema nesse horário, com as mesmas premissas de demanda (volume e TMA) e com a mesma capacidade efetiva dimensionada (12 agentes).

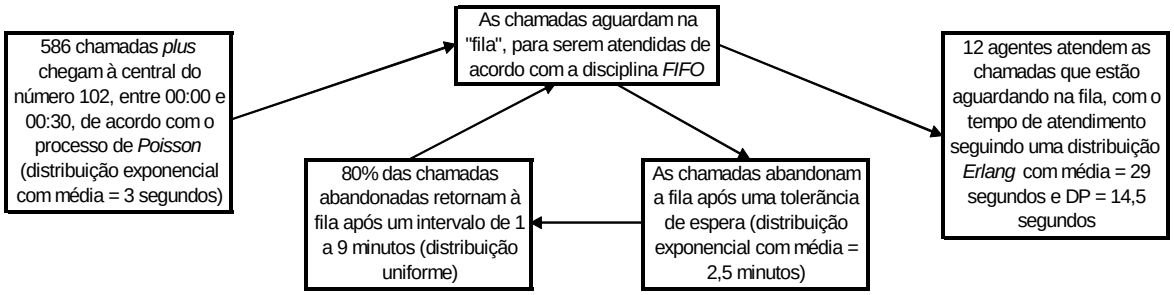
Como as chamadas vão chegando à central sem nenhuma espécie de controle, este processo pode ser considerado aleatório, caso em que a base conceitual sugere que a taxa de chegada de ligações seja modelada através de um processo de Poisson. O modelo de simulação construído implementou este processo com média de 0,33 (= 586 / (30 minutos x 60 segundos)) ligações chegando por segundo (ou seja, com uma chamada chegando a cada 3 segundos, aproximadamente).

Em relação ao tempo de atendimento, a distribuição Erlang costuma modelar bem esse processo e, portanto, foi utilizada, com média de 29 segundos. Ela requer, no entanto, um parâmetro adicional (k) que diz respeito à variabilidade dos dados em torno da média. O desvio-padrão da distribuição é igual à sua média dividida pela raiz quadrada de k. Para, então, considerar uma variabilidade moderada dos dados em torno da média, o modelo utiliza a distribuição Erlang com $k = 4$, fazendo com que o coeficiente de variação (desvio-padrão dividido pela média) seja de 50%.

Para que o modelo contemplasse corretamente o comportamento de abandono dos clientes, foi necessário realizar uma pesquisa junto à base de dados da Contax, contendo as ligações abandonadas, para o produto 102. Ela mostrou que o tempo de espera das ligações abandonadas apresenta, historicamente, uma média em torno de 2,5 minutos, seguindo uma distribuição não muito distante de uma exponencial. Também foi preciso modelar o comportamento de retorno das ligações abandonadas. Para tal, foi utilizada a premissa de que 80% das ligações abandonadas são refeitas posteriormente, entre 1 e 9 minutos após o abandono (distribuição uniforme).

De forma esquemática, o modelo de simulação pode ser descrito e visualizado através da figura 36 a seguir.

Figura 36 – Diagrama de processos do modelo básico de simulação, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes

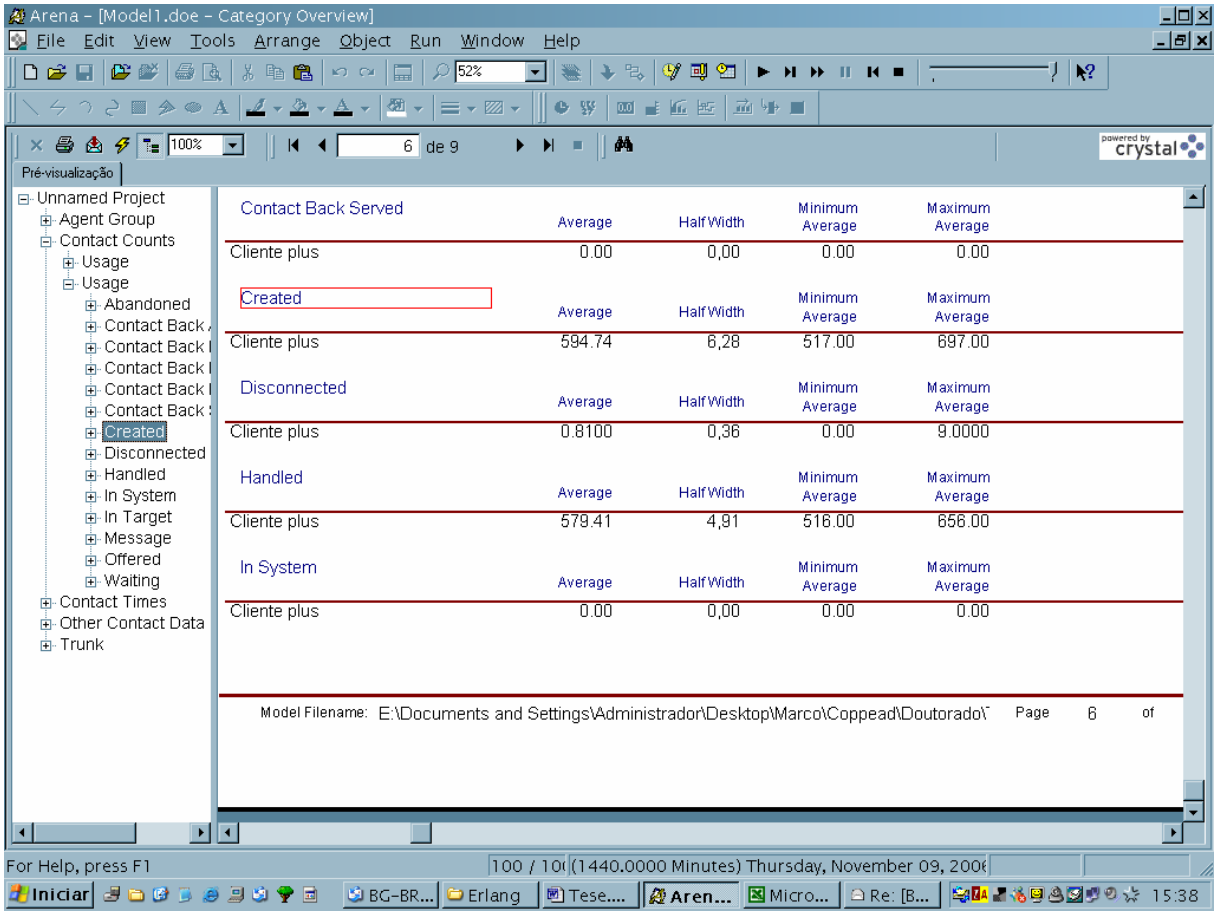


Fonte: Diagrama elaborado pelo pesquisador

A simulação dos 30 minutos de operação do *call center* foi replicada 100 vezes no *software*, em 142 segundos. O primeiro resultado aparece na figura 37 a seguir.

Em média, 595 chamadas foram geradas (*created*) em cada replicação. Este número é um pouco maior do que a premissa de demanda (586 chamadas) porque, na simulação, algumas das chamadas abandonadas voltaram a ser geradas e entraram novamente na fila. Das chamadas geradas, 579 – em média – foram efetivamente atendidas (*handled*) pelos agentes em cada replicação.

Figura 37 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

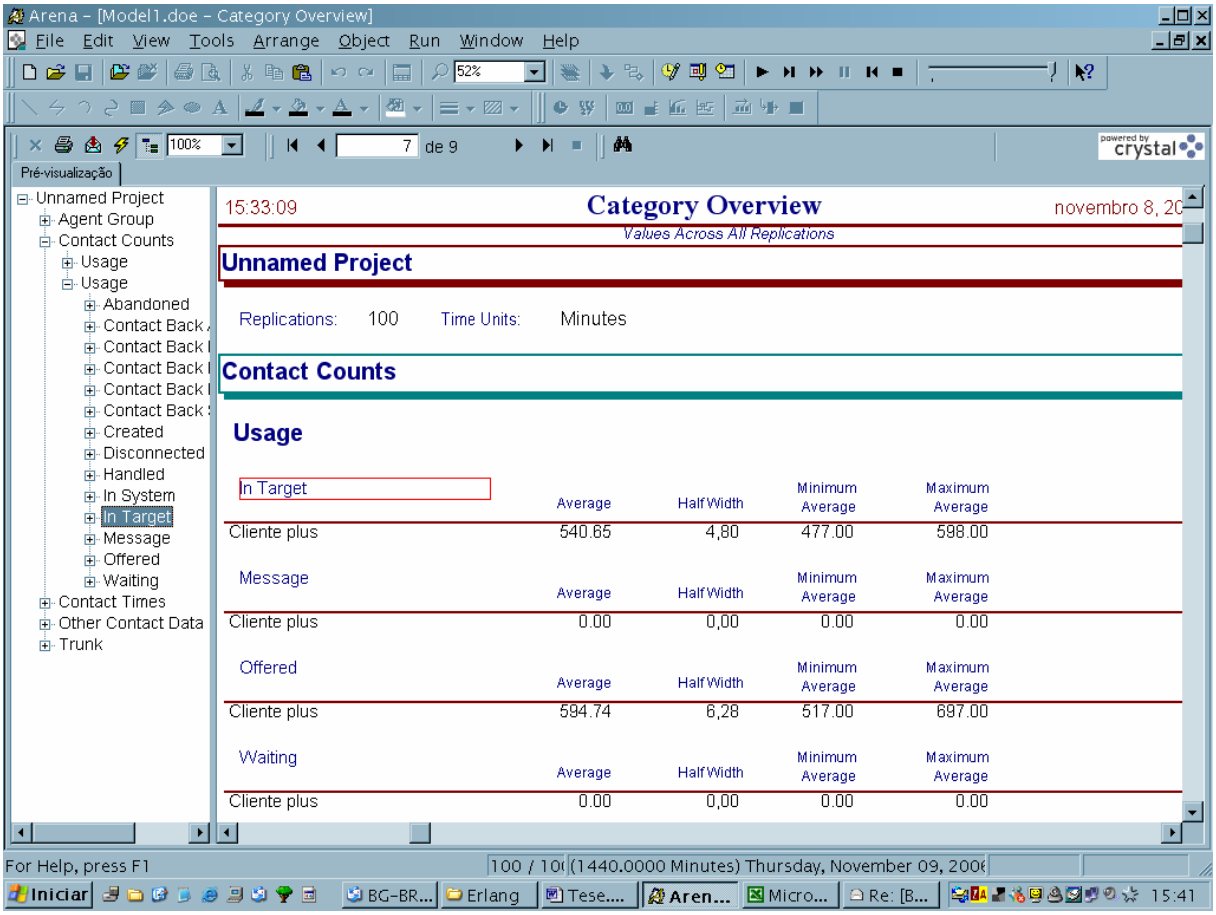
O percentual destas chamadas que foi atendido em até 10 segundos (nível de serviço) pode ser calculado através de uma informação extraída da figura 38 a seguir.

Em média, 541 chamadas foram atendidas em até 10 segundos (*in target*). O nível de serviço obtido da divisão desse valor pelo total de chamadas atendidas (579) foi 93,31%. Esse valor é razoavelmente maior do que os 88,04% previstos de nível de serviço segundo a abordagem analítica. Esse fato confirma, em princípio, a expectativa de subestimação desta grandeza por parte da teoria de filas.

Depois de realizada uma consulta junto à base de dados da Contax, foi possível recuperar a informação de que, no dia 22 de agosto de 2006, estiveram presentes 12 operadores durante o

período das 00:00 às 00:30. Dentro deste intervalo, 592 ligações foram atendidas, em 29,4 segundos cada, em média.

Figura 38 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações



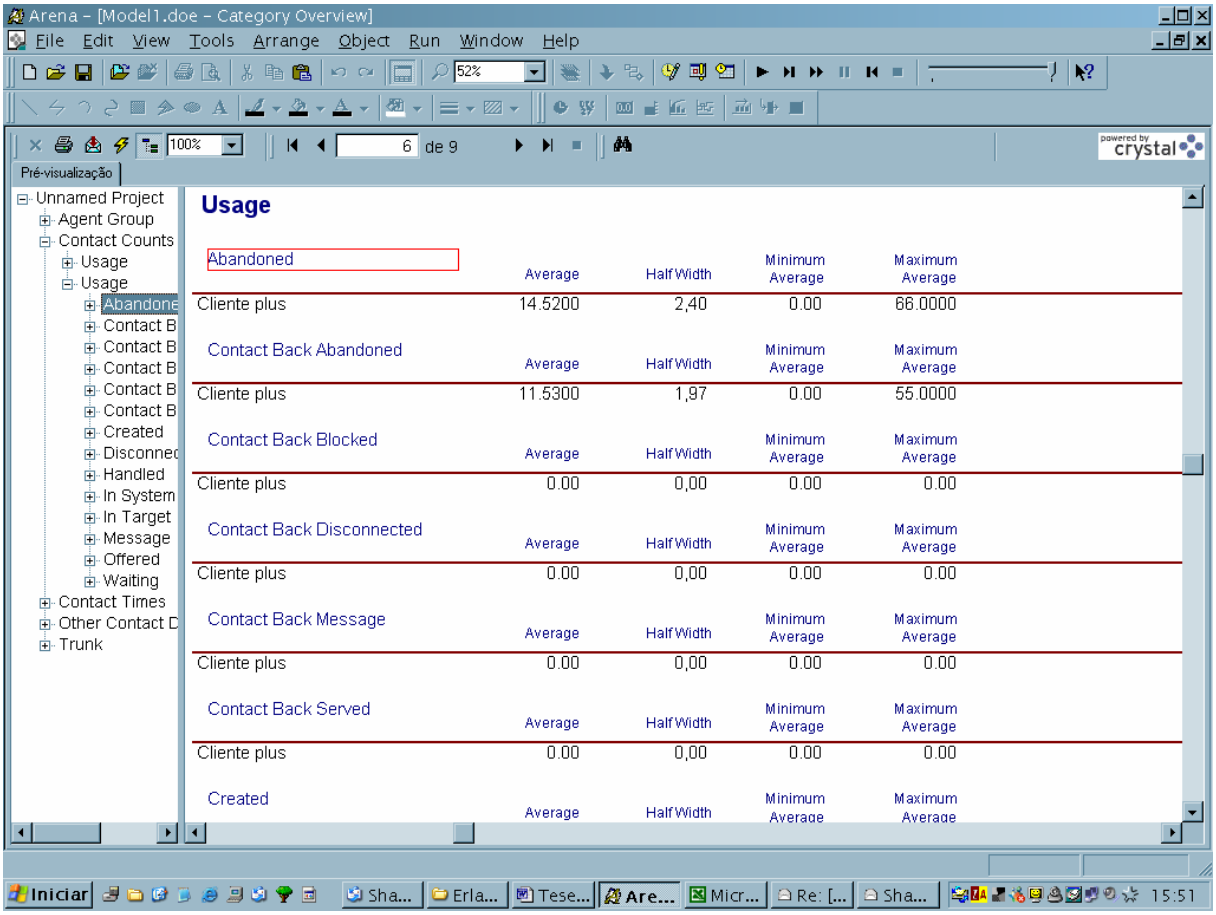
Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Estes números superam por muito pouco as premissas de demanda (volume = 586; TMA = 29 segundos) utilizadas no dimensionamento via fórmulas analíticas e também no modelo de simulação aqui apresentado. Por isso, esse dia e horário foram escolhidos para servirem de base de comparação entre os resultados obtidos pelas duas abordagens.

Durante estes 30 minutos, 549 das 592 chamadas foram atendidas em até 10 segundos; ou seja, o nível de serviço real ocorrido neste intervalo foi de 92,74%, bem mais próximo do valor calculado empiricamente pela simulação (93,31%) do que do valor estimado analiticamente pelas fórmulas Erlang (88,04%).

Conforme foi dito, a subestimação do nível de serviço promovida pela teoria de filas deve-se, sobretudo, à não contemplação do abandono da ligação, empreendido – na prática – por uma parcela dos clientes, mas desconsiderado – na teoria – pelos modelos analíticos. O comportamento do abandono durante a simulação pode ser visualizado na figura 39 a seguir.

Figura 39 – Chamadas abandonadas e retornadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações



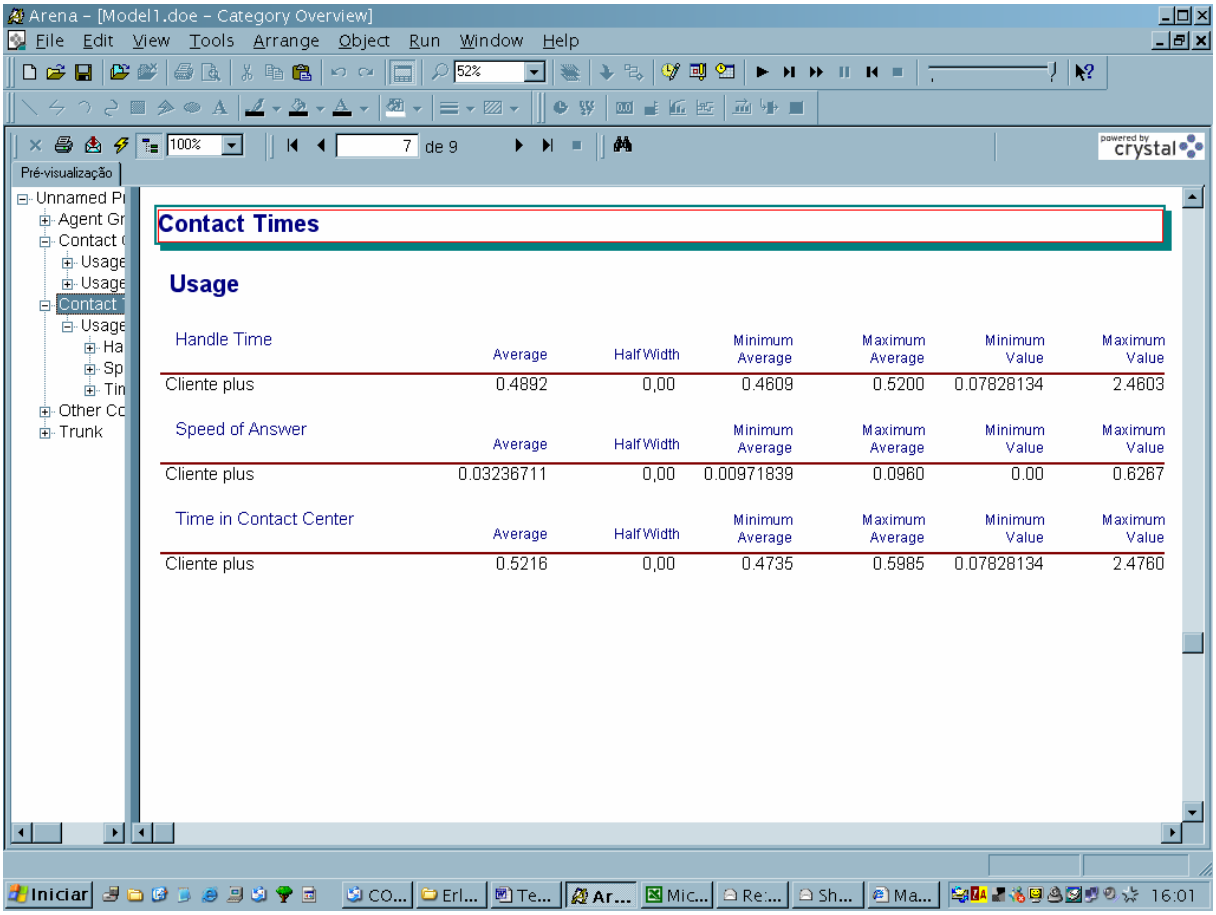
Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Das 595 chamadas geradas em cada replicação, 14,5 – em média – foram abandonadas (*abandoned*) pelos clientes, ocasionando uma taxa de abandono igual a 2,44%. Dentre as 14,5 chamadas abandonadas, 11,5 (79,41%) retornaram para a fila (*contact back abandoned*) alguns minutos após o abandono.

Como deve estar sendo observado, a simulação permite que muitos outros indicadores de performance – além do nível de serviço – a respeito das chamadas possam ser observados e

analisados. A figura 40 a seguir apresenta os tempos de atendimento e de espera experimentados pelas chamadas.

Figura 40 – Tempos de atendimento e tempos de espera, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações



Contact Times

Usage

Handle Time	Average	HalfWidth	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Cliente plus	0.4892	0,00	0.4609	0.5200	0.07828134	2.4603

Speed of Answer	Average	HalfWidth	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Cliente plus	0.03236711	0,00	0.00971839	0.0960	0.00	0.6267

Time in Contact Center	Average	HalfWidth	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Cliente plus	0.5216	0,00	0.4735	0.5985	0.07828134	2.4760

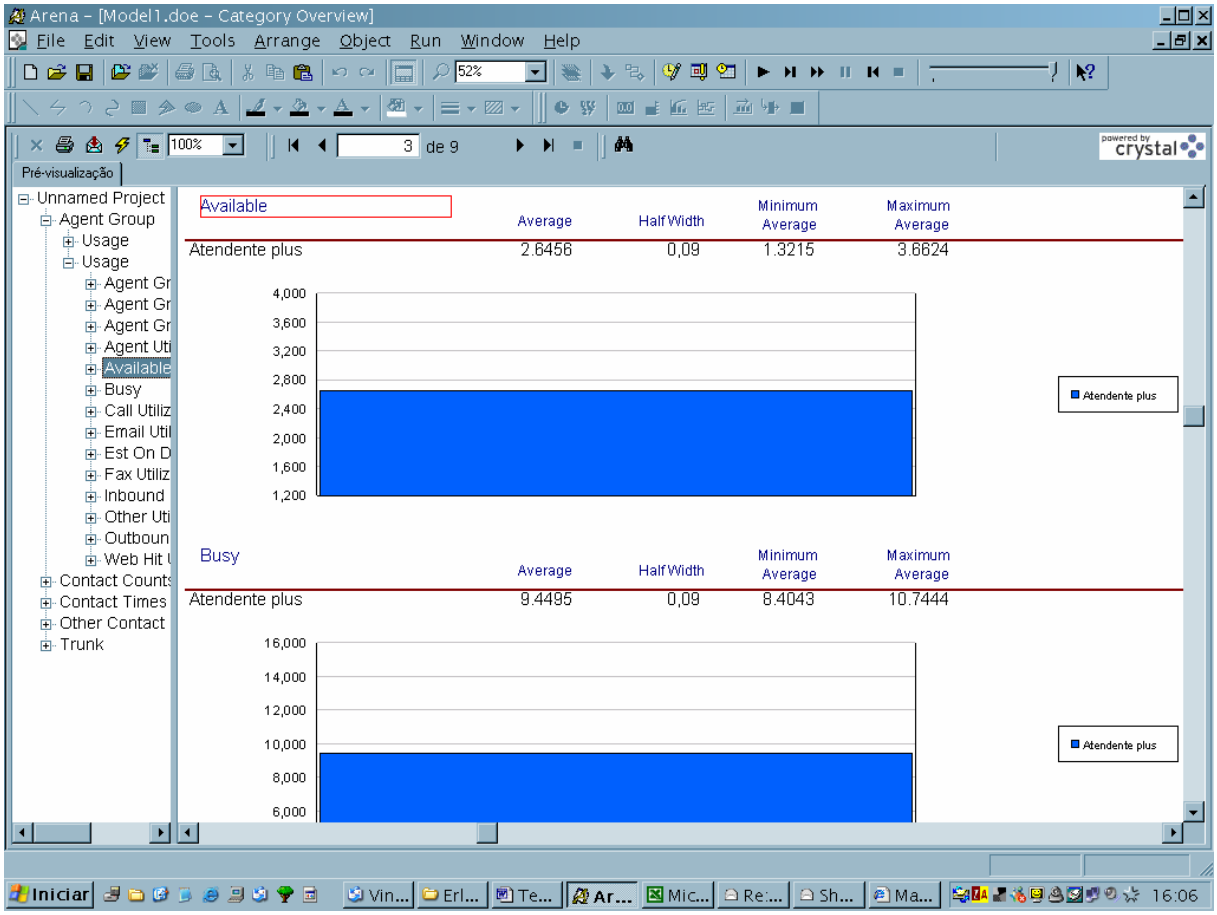
Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

O tempo médio de atendimento (*handle time*) foi de 0,49 minutos (ou 29,35 segundos; bem próximo da premissa de 29 segundos usada para o TMA). Mas houve uma grande variabilidade nesse tempo, já que um cliente chegou a ser atendido em 0,0783 minutos (*minimum value*) e outro em 2,46 minutos (*maximum value*), aproximadamente. Já em relação ao tempo de espera (*speed of answer*), as chamadas aguardaram, em média, 0,032 minutos (ou 1,94 segundos) para serem atendidos. Alguns clientes foram atendidos instantaneamente (*minimum value*) e pelo menos um esperou 0,63 minutos (*maximum value*) – ou quase 38 segundos – antes de ser atendido.

Conforme alertado na seção 3.5, esse tipo de análise envolvendo a variabilidade e os valores máximos e mínimos dos indicadores de desempenho não pode ser realizado por meio de métodos analíticos (capazes de apresentar apenas os valores médios), tornando-se viável apenas através de abordagens experimentais.

Naturalmente, a simulação também gera indicadores a respeito dos agentes, conforme ilustrado na figura 41 a seguir.

Figura 41 – Agentes disponíveis e ocupados, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Dos 12 atendentes presentes no horário (13 teriam sido designados, mas o absenteísmo teria reduzido esse efetivo), 2,65 ficaram disponíveis (*available*), em média, ao longo dos 30 minutos; e 9,45 estiveram ocupados (*busy*), em média, durante o mesmo período. Esses números podem ser convertidos na taxa média de ocupação dos agentes, durante esse intervalo de tempo: 78,75% (= 9,45 / 12). Este indicador pode orientar a gerência no sentido

de um aumento ou redução de efetivo, de acordo com as metas de ocupação traçadas *a priori*, conforme será ilustrado ainda nesta seção.

Todos estes resultados partiram do pressuposto que o tempo de atendimento segue uma distribuição Erlang (com k = 4 e, conseqüentemente, coeficiente de variação = 50%). No entanto, é bem possível que o tempo de atendimento apresente uma variabilidade (desvio-padrão) diferente (maior ou menor do que metade da média), e que este parâmetro seja capaz de impactar os resultados mais importantes.

A análise de sensibilidade acerca da variabilidade do tempo de atendimento – cujos resultados estão apresentados na tabela 37 a seguir – busca mensurar esse impacto. A mesma simulação foi repetida algumas vezes no *software*, sempre com a mesma média para o tempo de atendimento (29 segundos), mas com diferentes valores de k (e, conseqüentemente, do coeficiente de variação e, em última análise, do desvio-padrão), de onde os *outputs* relevantes foram coletados, a partir dos quais os principais indicadores de desempenho (nível de serviço – percentual de ligações atendidas em até 10 segundos – e taxa de abandono) puderam ser obtidos, estando apresentados na tabela 37 a seguir.

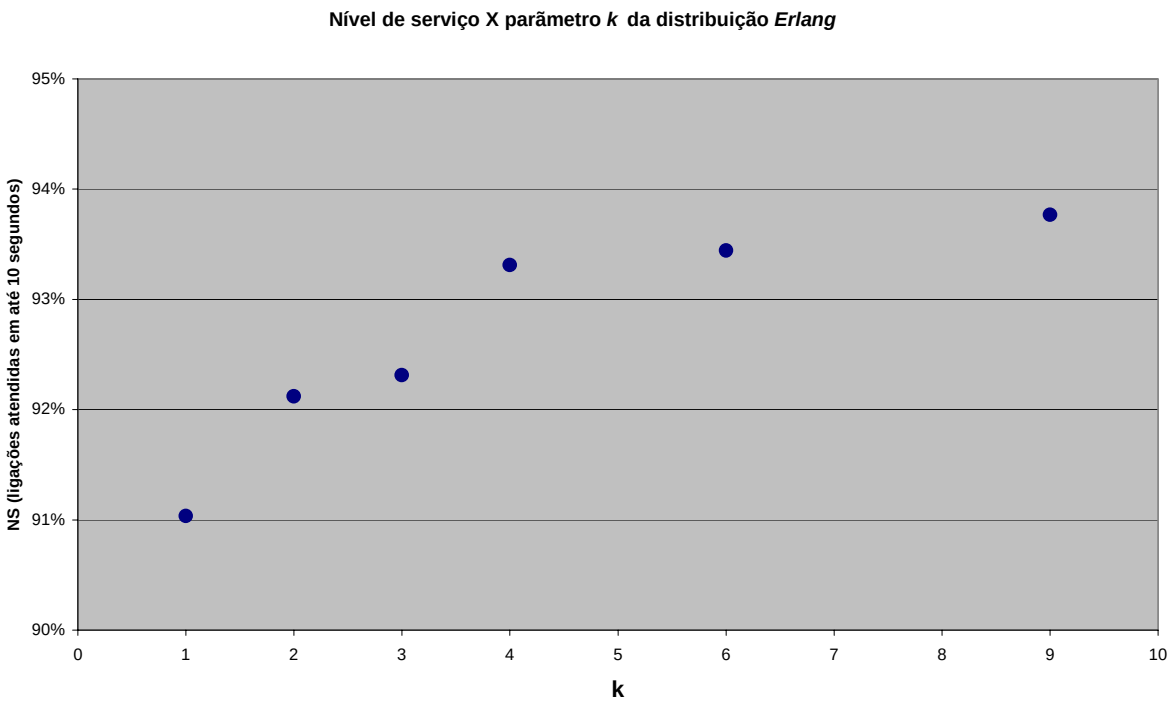
Tabela 37 – Nível de serviço e taxa de abandono para diferentes valores do parâmetro *k* da distribuição Erlang, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações

k	Coeficiente de variação	Ligações				Nível de serviço	Taxa de abandono
		geradas	atendidas	em até 10 seg	abandonadas		
1	100%	597	580	528	16,17	91,03%	2,71%
2	71%	602	585	539	16,09	92,12%	2,67%
3	58%	598	584	539	13,27	92,31%	2,22%
4	50%	595	579	541	14,52	93,31%	2,44%
6	41%	599	583	545	15,46	93,44%	2,58%
9	33%	597	583	546	13,92	93,77%	2,33%

Fonte: Tabela elaborada pelo pesquisador a partir de resultados obtidos pelo *software* Arena Contact Center

À medida que o valor de k aumenta, a variabilidade dos tempos de atendimento diminui. Devido à homogeneidade destes tempos, o sistema fica mais estável, apresentando como conseqüência mais notável o aumento do nível de serviço. Essa tendência evidenciada na tabela 37 anterior pode ser mais bem visualizada através do gráfico apresentado na figura 42 a seguir.

Figura 42 – Nível de serviço para diferentes valores do parâmetro *k* da distribuição Erlang, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações

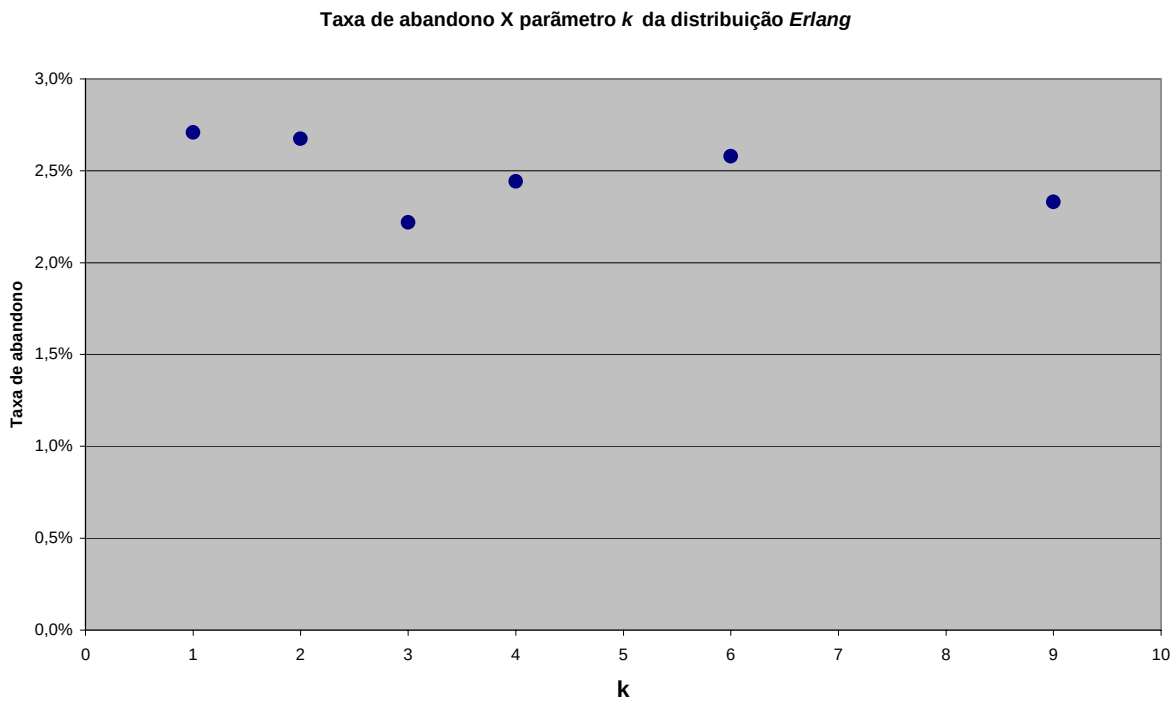


Fonte: Gráfico elaborado pelo pesquisador

Já a taxa de abandono não apresenta a mesma clareza na sua tendência, embora pareça cair à medida que a variabilidade do tempo de atendimento diminui (*k* aumenta), conforme pode ser observado no gráfico apresentado na figura 43 a seguir.

Em suma, à medida que a variabilidade do tempo de atendimento aumenta (*k* diminui), o sistema fica mais instável e, de uma maneira geral, seus indicadores de desempenho pioram: o nível de serviço (principalmente) diminui e a taxa de abandono aumenta. As variações nestes *outputs* não são imensas, mas estão longe de serem desprezíveis, revelando um significativo impacto potencial deste parâmetro nos resultados mais importantes. A correta consideração, portanto, da variabilidade do tempo de atendimento – e não só a da sua média – revela-se extremamente necessária para a obtenção de resultados acurados.

Figura 43 – Taxa de abandono para diferentes valores do parâmetro *k* da distribuição Erlang, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações



Fonte: Gráfico elaborado pelo pesquisador

O pior desempenho (nível de serviço = 91,03% e taxa de abandono = 2,71%) ocorreu na situação em que a variabilidade é a maior possível (k = 1; coeficiente de variação = 100%). Este é um caso particular da distribuição Erlang em que ela coincide com a distribuição exponencial, justamente o formato utilizado para modelar o tempo de atendimento na metodologia analítica empreendida pela Contax para estimar os indicadores. Ou seja, a utilização de um formato de distribuição (exponencial) para o tempo de atendimento menos condizente com a realidade – e não só a desconsideração do comportamento de abandono do cliente – pode estar fazendo com que a abordagem empregada pela empresa subestime os indicadores de desempenho do sistema e obrigue o dimensionamento a alocar mais recursos, de forma desnecessária.

Essa constatação faz despertar a curiosidade a respeito da verificação dos resultados de uma simulação considerando um outro tipo de distribuição – já que, de acordo com o que foi levantado na seção 3.5, o comportamento do tempo de atendimento em *call centers* pode

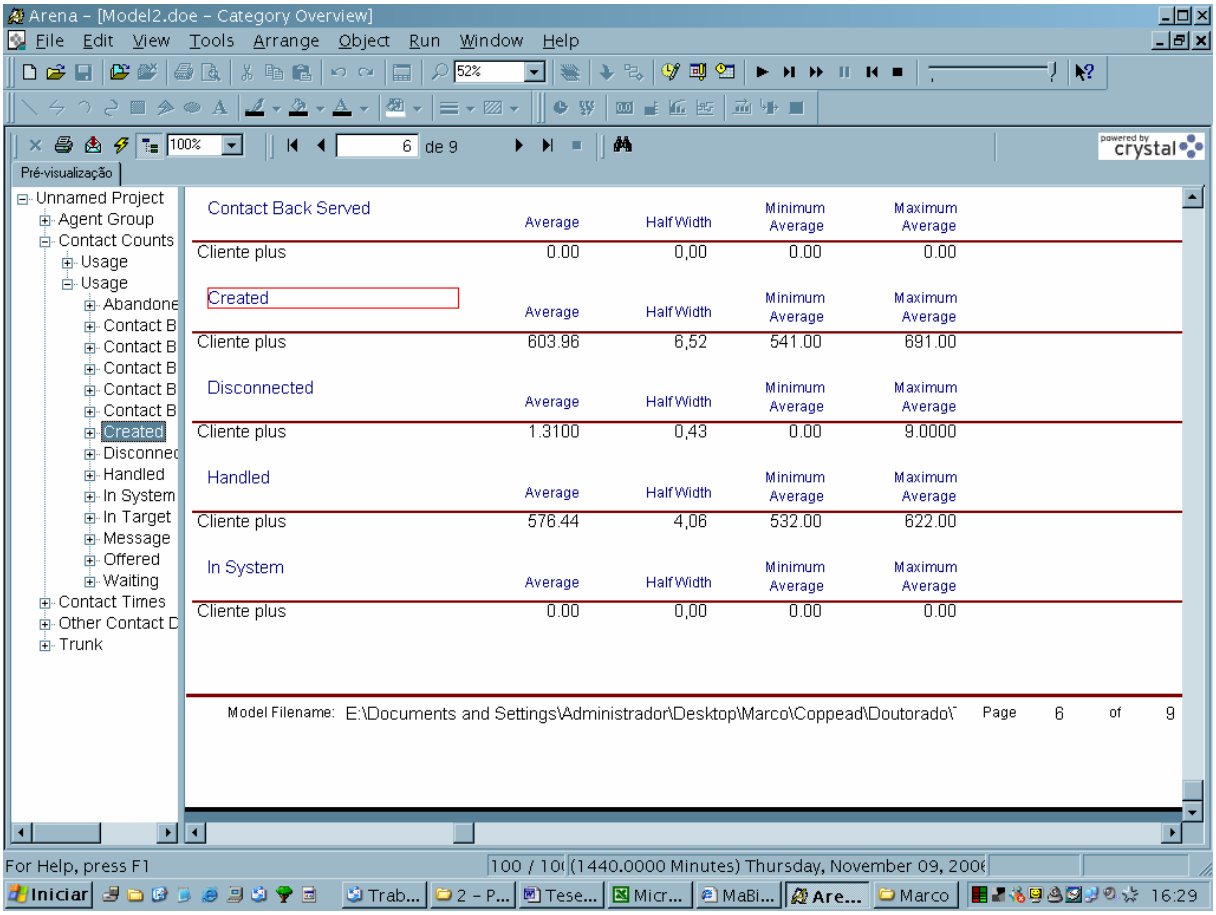
apresentar diferentes formatos – também comumente utilizado para modelar esta variável: a lognormal. Fazendo uso do mesmo modelo de simulação original, mas modificando a distribuição desta variável para o formato em questão (com a mesma média – 29 segundos e também mantendo o desvio-padrão em metade desse valor), 100 replicações foram realizadas no *software* Arena Contact Center.

Em média, 597 ligações foram geradas, 583 atendidas (sendo 544 em até 10 segundos) e 13, 95 abandonadas. O nível de serviço e a taxa de abandono resultantes foram 93,40% e 2,34%, respectivamente. Estes indicadores são muito próximos dos que foram obtidos com a distribuição Erlang com k = 4 (93,31% e 2,44%, respectivamente), conferindo uma certa confiabilidade a estes últimos e sugerindo que qualquer dos dois formatos comumente utilizados para modelar o tempo de atendimento possa ser utilizado indistintamente.

A simulação também permite a análise de cenários (*What-if*?). No exemplo que está sendo estudado, já que o emprego de 12 atendentes efetivos (descontado o absenteísmo) no horário em questão gerou um nível de serviço (93,31%) razoavelmente superior à meta mínima (85%), o que aconteceria com esse indicador após a redução de 1 atendente (para 11 efetivos)? Será que ele ainda se manteria acima da meta?

As figuras 44 e 45 a seguir apresentam os indicadores para o cálculo do nível de serviço em um cenário com 1 atendente a menos.

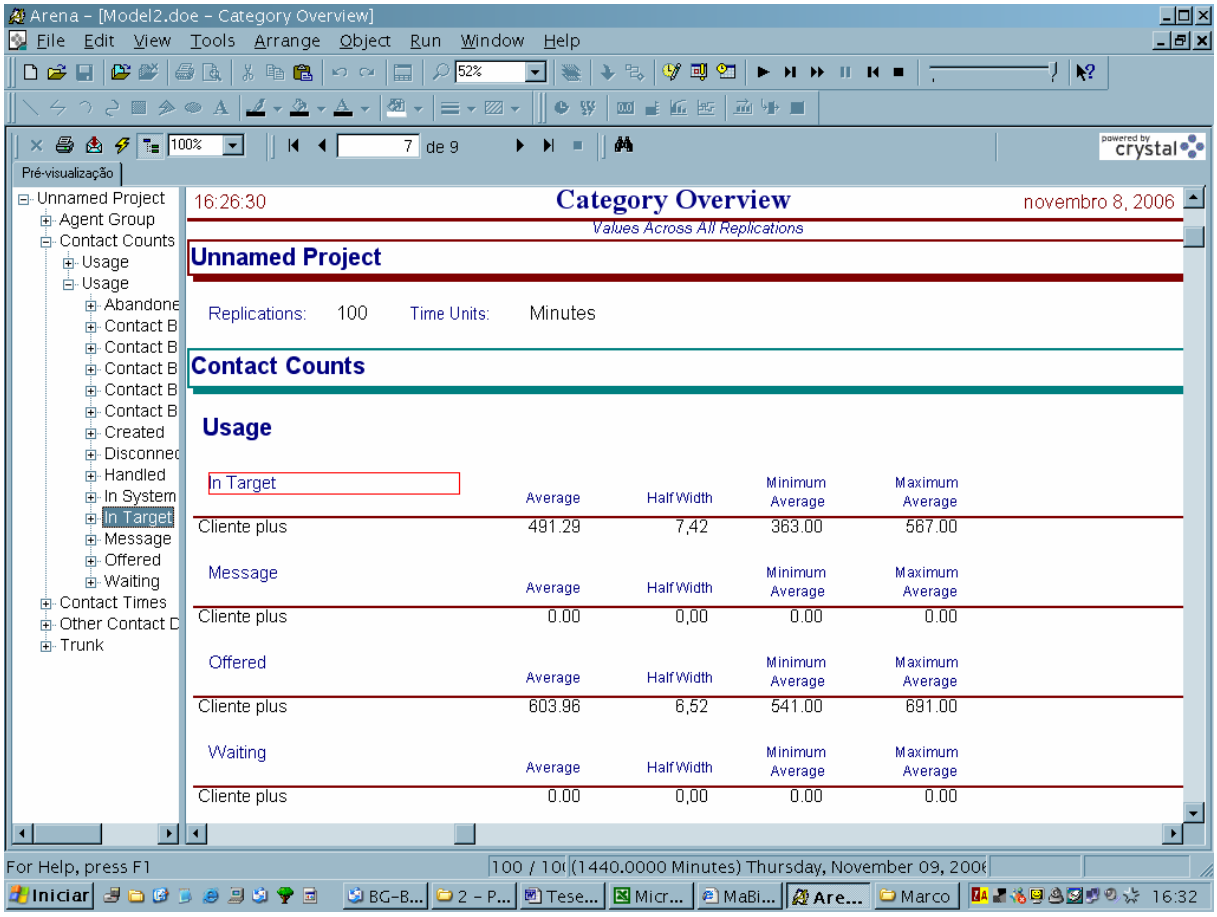
Figura 44 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 11 atendentes, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Dentro deste cenário, 576 chamadas foram atendidas, em média; das quais, 491 – em média – em até 10 segundos. O nível de serviço obtido no cenário com 11 atendentes foi, então, de 85,23%. Esse valor ainda está um pouco acima da meta de 85% estabelecida para os clientes *plus*.

Figura 45 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 11 atendentes, 100 replicações

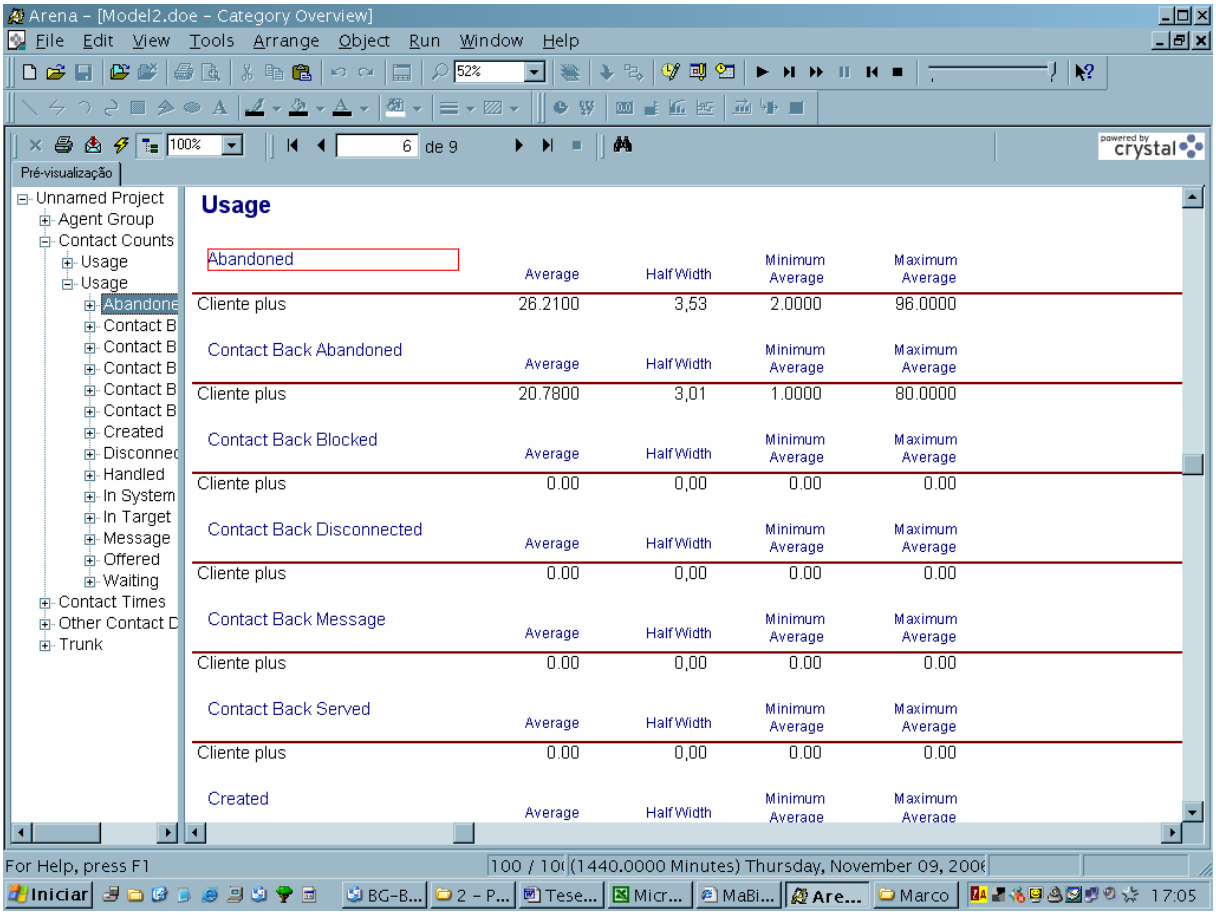


Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Em outras palavras, o 12º agente não fez falta (em termos da obtenção da meta de nível de serviço), apesar da sua ausência ter derrubado o nível de serviço em mais de 8 pontos percentuais. Mas, pode ser interessante também querer saber o impacto desta ausência na taxa de abandono, que pode ser observado na figura 46 a seguir.

Das 604 chamadas geradas em cada replicação, 26,2 – em média – foram abandonadas pelos clientes, ocasionando uma taxa de abandono igual a 4,34%, revelando um impacto grande neste indicador de performance.

Figura 46 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 11 atendentes, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Em suma, o uso da simulação neste caso permitiu verificar que a presença do 12º agente não era necessária para que a meta de 85% de nível de serviço fosse atingida, contradizendo o que afirmava a abordagem analítica neste caso. Isso aconteceu – conforme a seção 3.5 sugere – principalmente devido à consideração do abandono por parte da abordagem experimental, mas pode também ter sido influenciado pela utilização de um formato de distribuição para o tempo de atendimento mais condizente com a realidade: Erlang (pela simulação) ao invés de exponencial (pelas fórmulas analíticas).

Mas, a ausência do 12º agente teria ocasionado um grande aumento na taxa de abandono, o que poderia comprometer esta política de recursos mais enxuta. Caso este indicador não fosse tão importante, no entanto, o uso da simulação poderia ter proporcionado a economia de 1

agente para este horário, o que pode efetivamente acontecer em algum cenário no qual a taxa de abandono encontre-se em um patamar mais baixo.

O impacto da redução de atendentes pode ser também visualizado no caso de maiores déficits de oferta de atendimento através de uma análise de sensibilidade mais completa. A mesma simulação foi repetida algumas vezes no *software*, sempre com os mesmos parâmetros, mas variando a quantidade de atendentes. Os *outputs* relevantes foram coletados, a partir dos quais os principais indicadores de desempenho (nível de serviço – percentual de ligações atendidas em até 10 segundos – e taxa de abandono) puderam ser obtidos, estando apresentados na tabela 38 a seguir.

Tabela 38 – Nível de serviço e taxa de abandono para diferentes quantidades de atendentes, horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 100 replicações

Atendentes	Ligações				Nível de serviço	Taxa de abandono
	geradas	atendidas	em até 10 seg	abandonadas		
9	719	540	160	168,02	29,64%	23,38%
10	639	567	354	67,68	62,46%	10,60%
11	604	576	491	26,21	85,23%	4,34%
12	**595**	**579**	**541**	**14,52**	**93,31%**	**2,44%**

Fonte: Tabela elaborada pelo pesquisador a partir de resultados obtidos pelo *software* Arena Contact Center

Como era totalmente previsível, à medida que a oferta de atendimento diminui, o serviço piora, assim como os seus indicadores de desempenho, principalmente a taxa de abandono, que quadruplica com a redução de 2 agentes. O nível de serviço está bem abaixo da meta com este contingente, mas ainda não pode ser considerado inaceitável, o que ocorre no cenário com 9 atendentes, quando ele fica 3 vezes menor que o valor original. Com esta oferta de atendimento, o número de ligações abandonadas chega a superar as atendidas em até 10 segundos, tornando a taxa de abandono quase tão alta quanto o nível de serviço!

Isso tudo revela um impacto muito grande da redução da quantidade de atendentes no desempenho do sistema (indicando a necessidade de a atividade de dimensionamento ser desenvolvida com muito cuidado); e sugere que alocar 10 agentes para este horário seria o mínimo indispensável, cenário para o qual os indicadores de desempenho seriam ruins, mas não catastróficos.

Infelizmente, uma análise de sensibilidade mais completa acerca do aumento da quantidade de atendentes não seria muito pertinente neste caso, já que o nível de serviço já se encontra em um patamar bem alto (93%), deixando pouco espaço para melhorias potenciais. Tal análise será realizada ainda neste seção, em um horário apresentando o nível de serviço em um patamar mais baixo.

Conforme foi descrito na seção 5.2.2, o dimensionamento da capacidade de atendimento para a operação 102 é feito de forma isolada para cada produto – básico e *plus* – por causa da prioridade do último sobre o primeiro.

Mas a simulação permite visualizar o que aconteceria com a operação em um cenário em que os diferentes clientes – do tipo básico e *plus* – fossem atendidos de forma agregada (tanto por agentes "básicos" como por agentes "*plus*"), mas mantendo a prioridade para os últimos (descaracterizando, assim, a disciplina de fila comumente utilizada nos modelos analíticos – FIFO).

Para o dimensionamento e escalonamento dos operadores para atender os clientes básicos do produto 102 durante o mês de agosto de 2006, foi usada a premissa de que chegariam à central 399 chamadas destes clientes, com um tempo médio de atendimento de 34 segundos, na primeira meia-hora do dia (00:00 às 00:30), para quando foram dimensionados 7 atendentes.

A idéia agora consiste em simular – com o objetivo de observar o comportamento do sistema – o cenário onde essas chamadas se agregariam às chamadas dos clientes *plus* desse mesmo horário (cujas premissas foram mencionadas anteriormente nesta seção), formando uma única fila. Os 12 atendentes "*plus*" e os 7 "básicos" estariam aptos a atender ambos os tipos de chamadas, mas com diferentes proficiências e prioridades, caracterizando o roteamento baseado em habilidades, mecanismo viabilizado apenas em abordagens experimentais, conforme explicado anteriormente na seção 3.5.

Mas o cliente básico não se comporta da mesma forma que o cliente *plus* e suas diferentes características precisam ser contempladas pelo modelo. Já foi mencionado que o seu tempo de atendimento é, em média, um pouco maior. Para efeitos do modelo, a mesma distribuição Erlang foi utilizada para este tempo, com coeficiente de variação igual a 50% ($k = 4$).

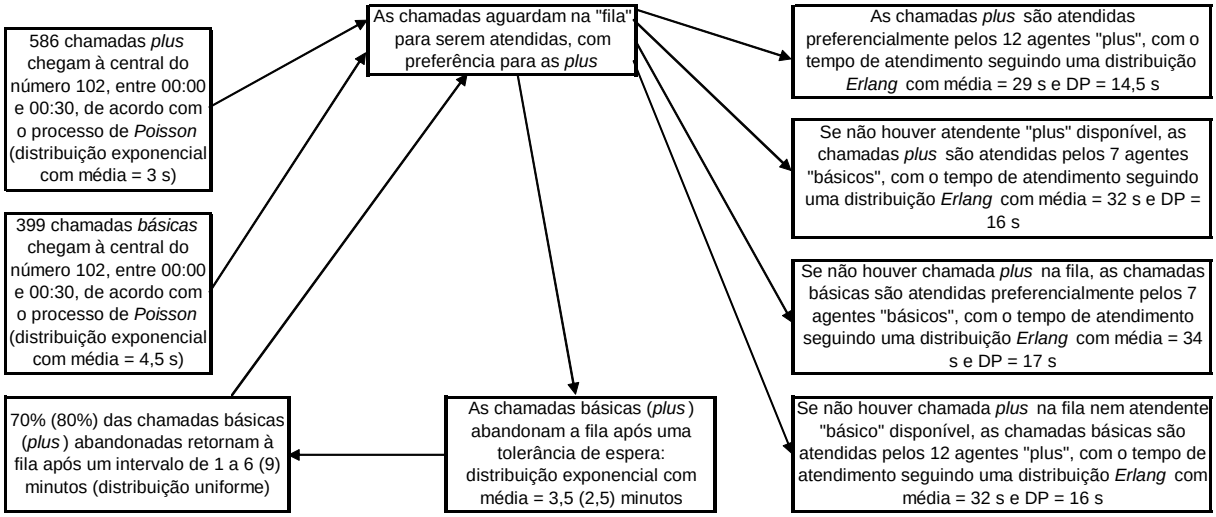
De uma forma geral, ele é mais paciente antes de abandonar a chamada e o tempo de espera antes do abandono também foi modelado como uma distribuição exponencial, mas com média de 3,5 minutos (ao invés de 2,5). Foi considerado também que uma parcela menor (70% ao invés de 80%) dos clientes que abandonam a ligação tentam refazê-la posteriormente, após um intervalo de tempo também menor: entre 1 e 6 minutos (distribuição uniforme).

Em termos de prioridades, as ligações *plus* têm preferência sobre as básicas ("furando a fila") e devem ser atendidas, sempre que possível, por atendentes *plus*, teoricamente mais capacitados; elas só seriam atendidas por agentes básicos no caso de todos os outros estarem ocupados. As ligações básicas, por sua vez, seriam atendidas preferencialmente por agentes básicos, teoricamente menos capacitados, de forma a deixar os "melhores" atendentes livres para as chamadas "mais importantes".

Quando as chamadas não são atendidas pelo seu tipo de agente preferencial, seu tempo de atendimento é modificado. O modelo considera que uma chamada *plus* sendo atendida por um agente básico (menos capacitado que o normal) demora 10% a mais para ser concluída; e que uma chamada básica sendo atendida por um agente *plus* (mais capacitado que o normal) demora 5% a menos para ser concluída.

De forma esquemática, o modelo de simulação pode ser descrito e visualizado através da figura 47 a seguir.

Figura 47 – Diagrama de processos do modelo de simulação com atendimento agregado (produto 102, básico e *plus*), das 00:00 às 00:30, agosto de 2006

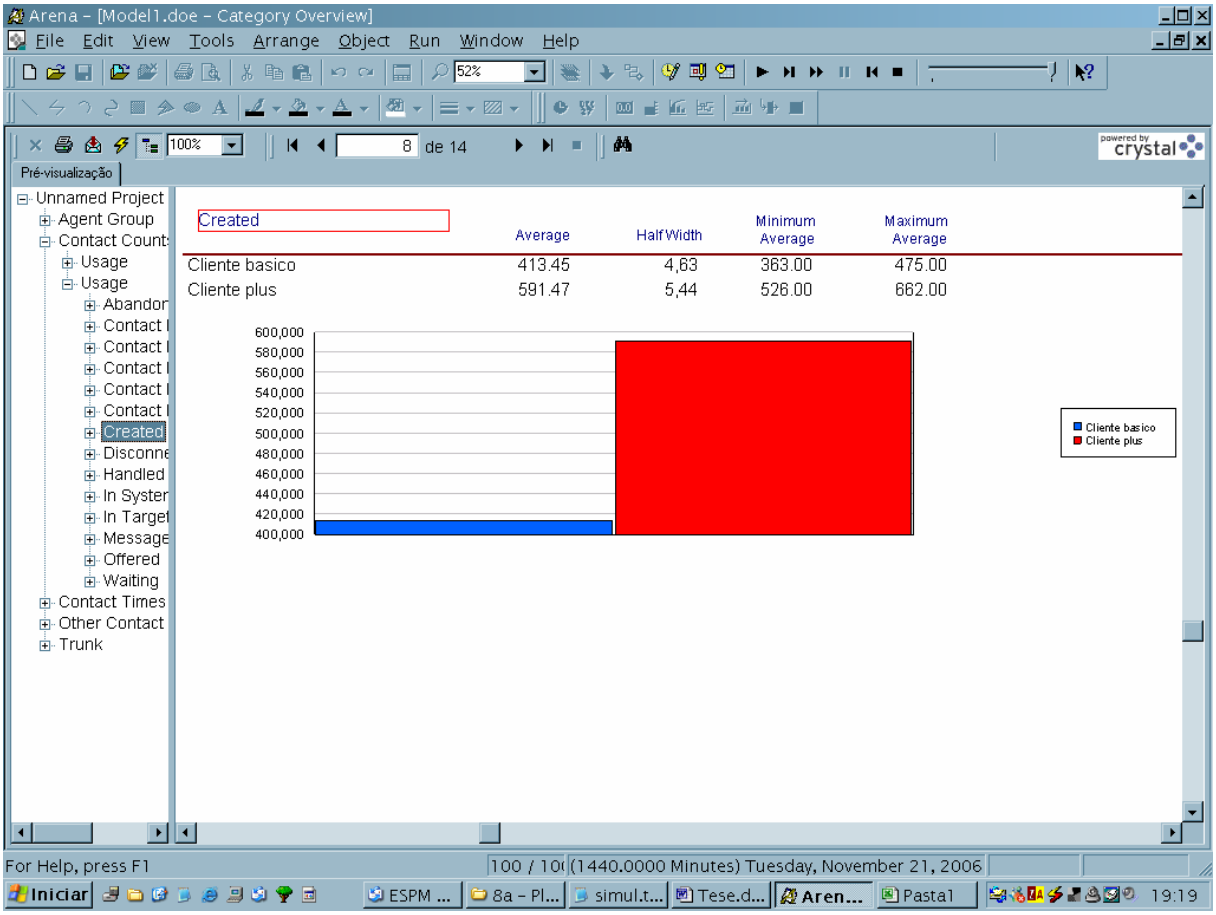


Fonte: Diagrama elaborado pelo pesquisador

A simulação foi replicada 100 vezes no *software* Arena Contact Center. Os resultados apresentados nas figuras 48 e 49 a seguir permitem analisar o comportamento de abandono.

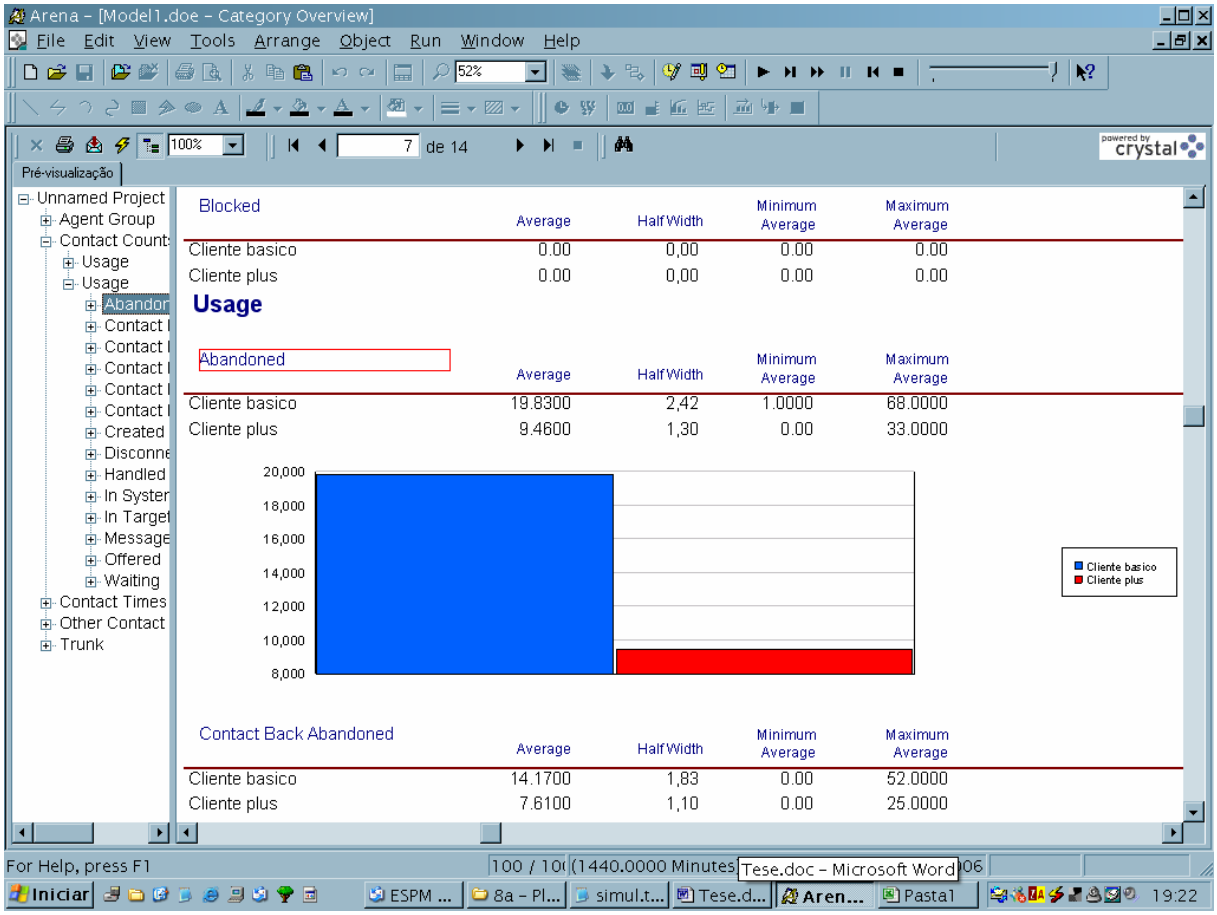
Em média, 413 chamadas básicas e 591 chamadas *plus* foram geradas em cada replicação, tendo sido abandonadas 19,8 e 9,5, respectivamente. As taxas de abandono resultantes foram 4,80% para os clientes básicos e 1,60% para os clientes *plus*. Dentre as ligações básicas abandonadas, 14,2 (71,46%) retornaram para a fila alguns minutos após o abandono; dentre as *plus*, 7,6 (80,44%) fizeram o mesmo.

Figura 48 – Chamadas geradas das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Figura 49 – Chamadas abandonadas e retornadas das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

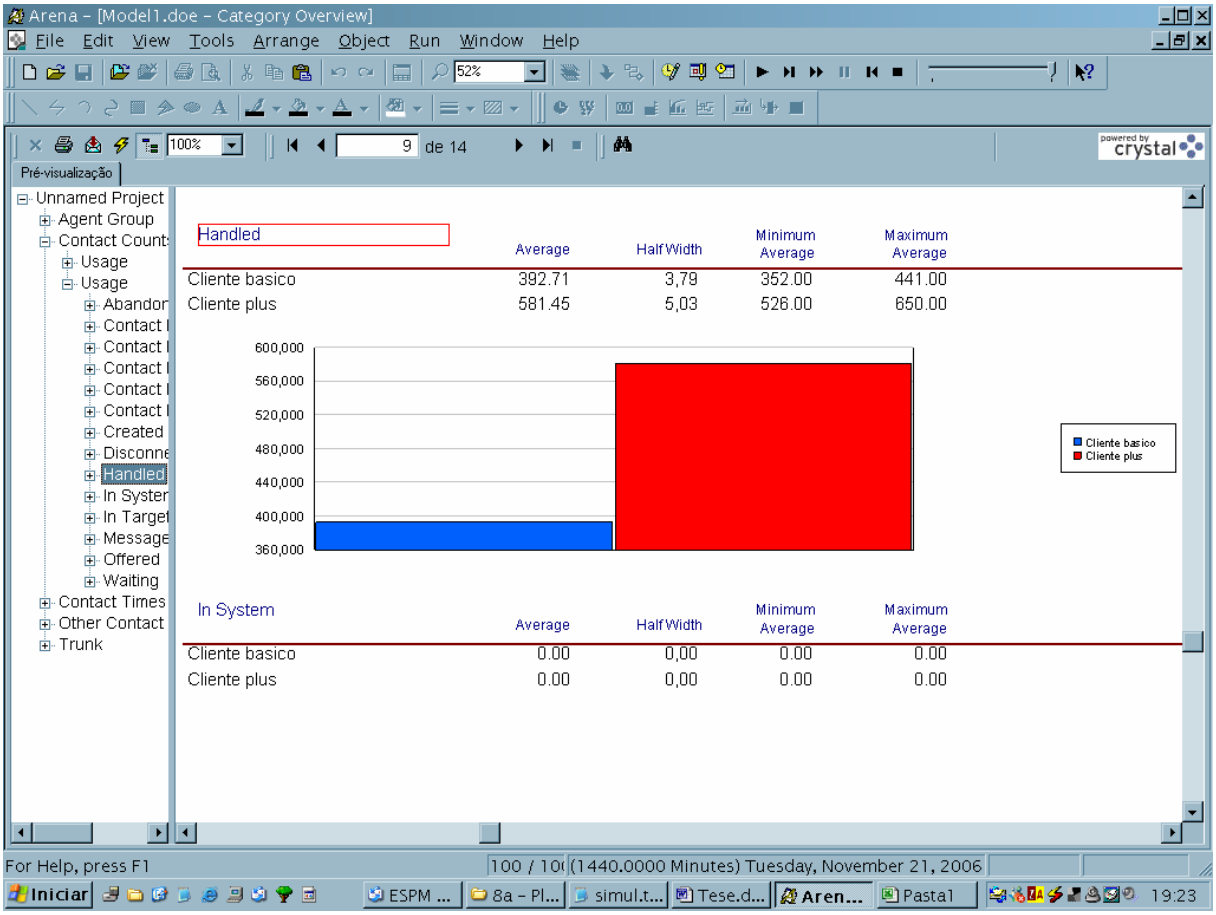
As figuras 50 e 51 a seguir contêm os insumos para a análise do nível de serviço.

393 chamadas básicas e 581 *plus* – em média – foram efetivamente atendidas pelos agentes em cada replicação. Destas, 303 básicas (77,25%) e 575 *plus* (98,96%) foram atendidas em até 10 segundos (nível de serviço).

Comparando os indicadores de desempenho dos clientes *plus* com os anteriormente obtidos no cenário de atendimento segmentado (taxa de abandono = 2,44% e nível de serviço = 93,31%), é fácil concluir que o atendimento agregado melhorou razoavelmente a operação, do ponto de vista destes clientes. Tal resultado era de se esperar, pois 7 agentes "básicos" passaram a atender as chamadas *plus*. É bem verdade que os 19 agentes (12 "*plus*" e 7

“básicos”) também atenderam as chamadas básicas, mas somente quando não havia uma chamada *plus* aguardando.

Figura 50 – Chamadas atendidas das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações



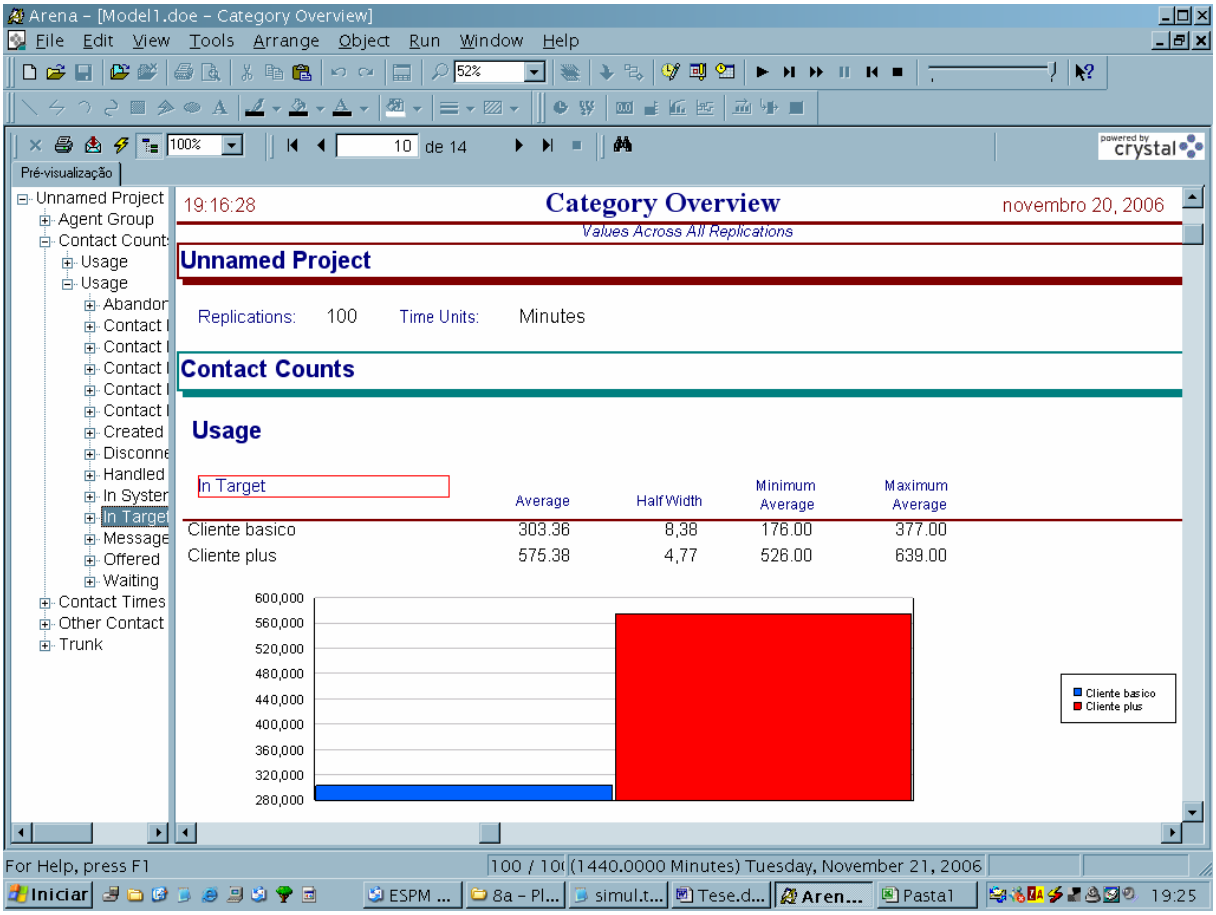
Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Esta preferência possibilitou a sensível melhoria no atendimento dos clientes *plus* e, certamente, reduziu um pouco a qualidade do serviço junto aos clientes básicos. Mas o interessante é notar que, neste cenário, o nível de serviço para os clientes preteridos (77,25%) manteve-se ainda acima da meta (75%). A taxa de abandono (4,80%) ficou um pouco alta, mas não chega a ser inaceitável para clientes básicos.

Em suma, o formato agregado do atendimento melhorou o desempenho do sistema para os clientes *plus* sem, no entanto, comprometer demais a qualidade do serviço para os clientes básicos. Isso aconteceu porque os agentes podiam atender as suas chamadas não preferenciais enquanto estavam ociosos, permitindo o incremento na qualidade da operação como um todo.

Análises e conclusões deste tipo não poderiam ser conduzidas nem obtidas através de metodologias analíticas, só sendo viáveis por meio de uma abordagem experimental, como a simulação.

Figura 51 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações

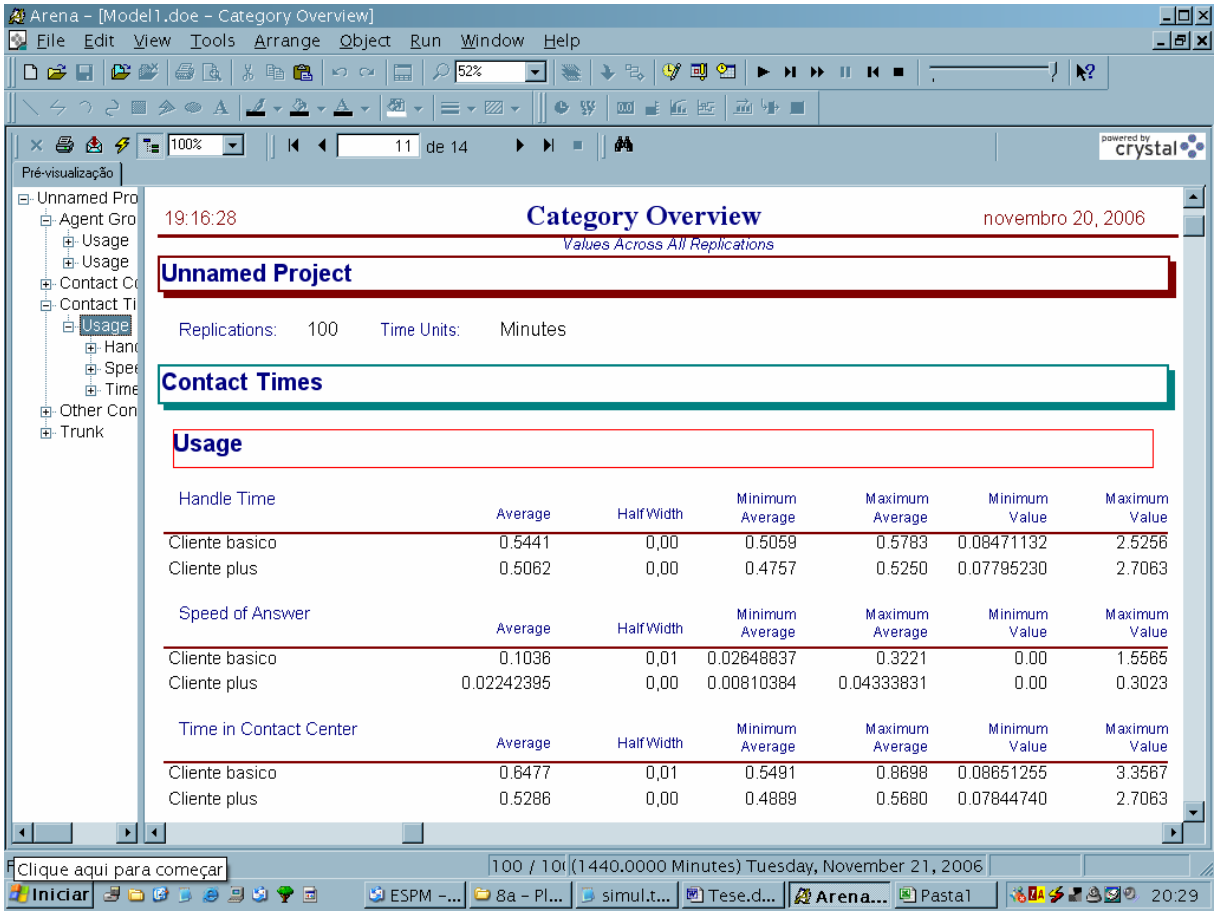


Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

A figura 52 a seguir apresenta outros indicadores interessantes de serem observados para este cenário.

O tempo médio de atendimento dos clientes básicos foi de 0,54 minutos (ou 32,65 segundos), um pouco inferior ao valor de 34 segundos, utilizado para a premissa do TMA, porque alguns deles foram atendidos por agentes mais rápidos (os "*plus*"). Para os clientes *plus*, o tempo médio de atendimento foi de 0,51 minutos (ou 30,37 segundos), um pouco superior à premissa de 29 segundos pelo fato de alguns deles terem sido atendidos por agentes mais lentos (os "básicos").

Figura 52 – Tempos de atendimento e tempos de espera, das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações



19:16:28 **Category Overview** novembro 20, 2006

Values Across All Replications

Unnamed Project

Replications: 100 Time Units: Minutes

Contact Times

Usage

Handle Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Cliente basico	0.5441	0,00	0.5059	0.5783	0.08471132	2.5256
Cliente plus	0.5062	0,00	0.4757	0.5250	0.07795230	2.7063

Speed of Answer	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Cliente basico	0.1036	0,01	0.02648837	0.3221	0.00	1.5565
Cliente plus	0.02242395	0,00	0.00810384	0.04333831	0.00	0.3023

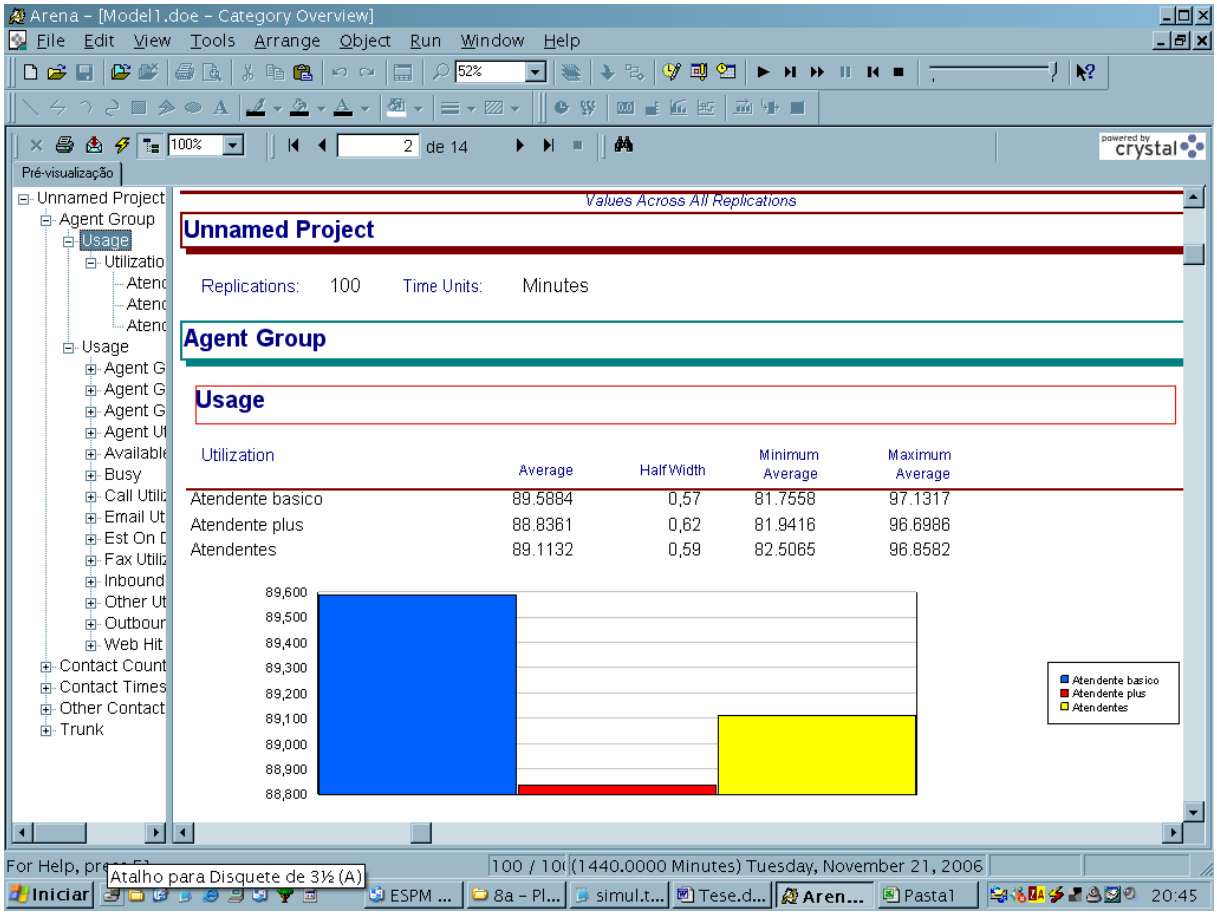
Time in Contact Center	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Cliente basico	0.6477	0,01	0.5491	0.8698	0.08651255	3.3567
Cliente plus	0.5286	0,00	0.4889	0.5680	0.07844740	2.7063

Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

As chamadas básicas aguardaram, em média, 0,104 minutos (ou 6,24 segundos) para serem atendidas, enquanto as chamadas *plus* apenas 0,022 minutos (ou 1,34 segundos). Tal disparate deveu-se à preferência de atendimento desfrutada por estas últimas.

A utilização dos agentes pode ser observada na figura 53 a seguir.

Figura 53 – Utilização dos agentes "básicos" e "*plus*", das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações

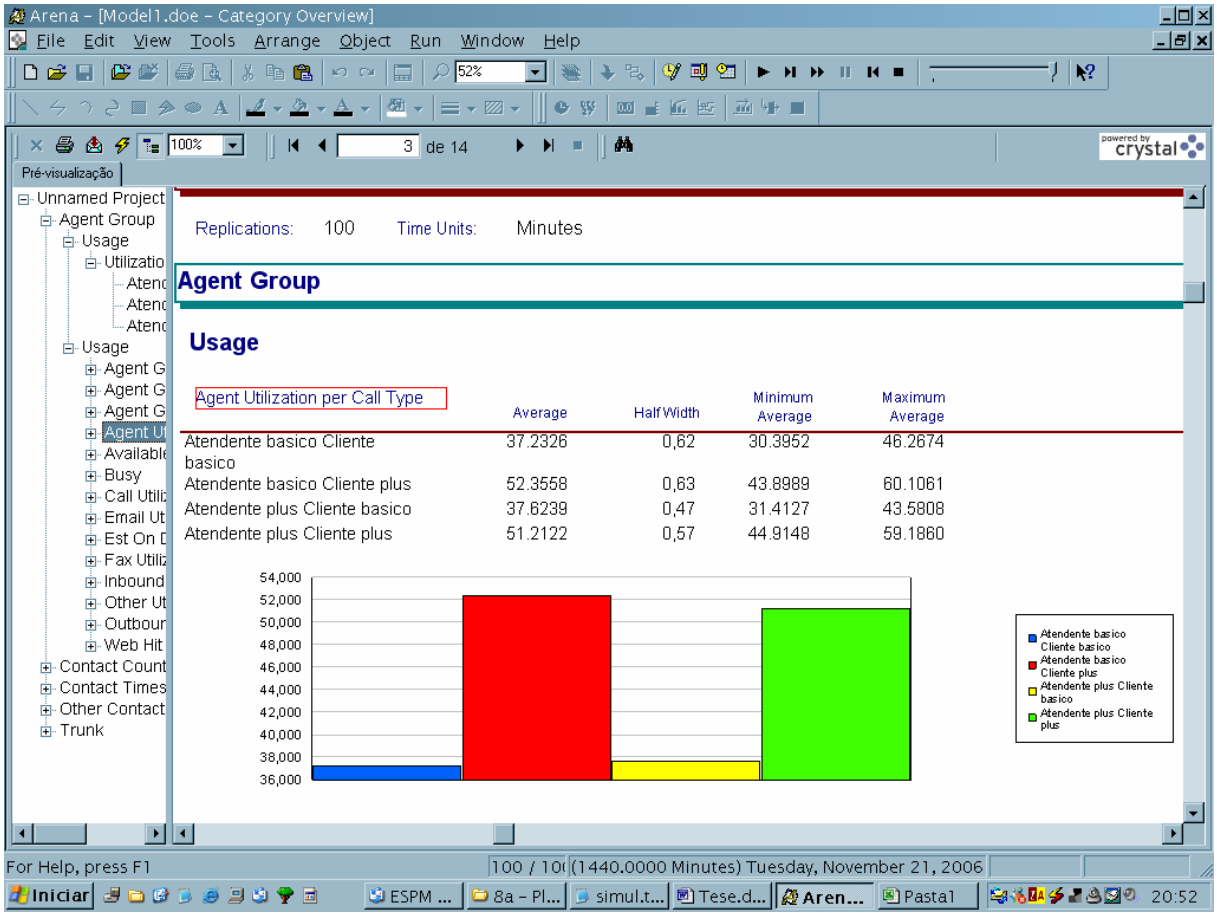


Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

A taxa média de ocupação dos atendentes "básicos" (89,59%) foi muito parecida com a dos atendentes "*plus*" (88,84%), principalmente pelo fato de ambos os tipos de agentes estarem aptos a atender ambos os tipos de chamadas. Isto pode ser visto através dos resultados apresentados na figura 54 a seguir.

É possível perceber que ambos os tipos de agentes estiveram a maior parte do tempo (51-52%) atendendo os clientes *plus* (preferenciais e mais numerosos) e o restante da jornada atendendo os clientes básicos (37-38%) ou ociosos.

Figura 54 – Utilização dos agentes "básicos" e "*plus*" por tipo de chamada, das 00:00 às 00:30, agosto de 2006, produto 102 (básico e *plus*), 100 replicações



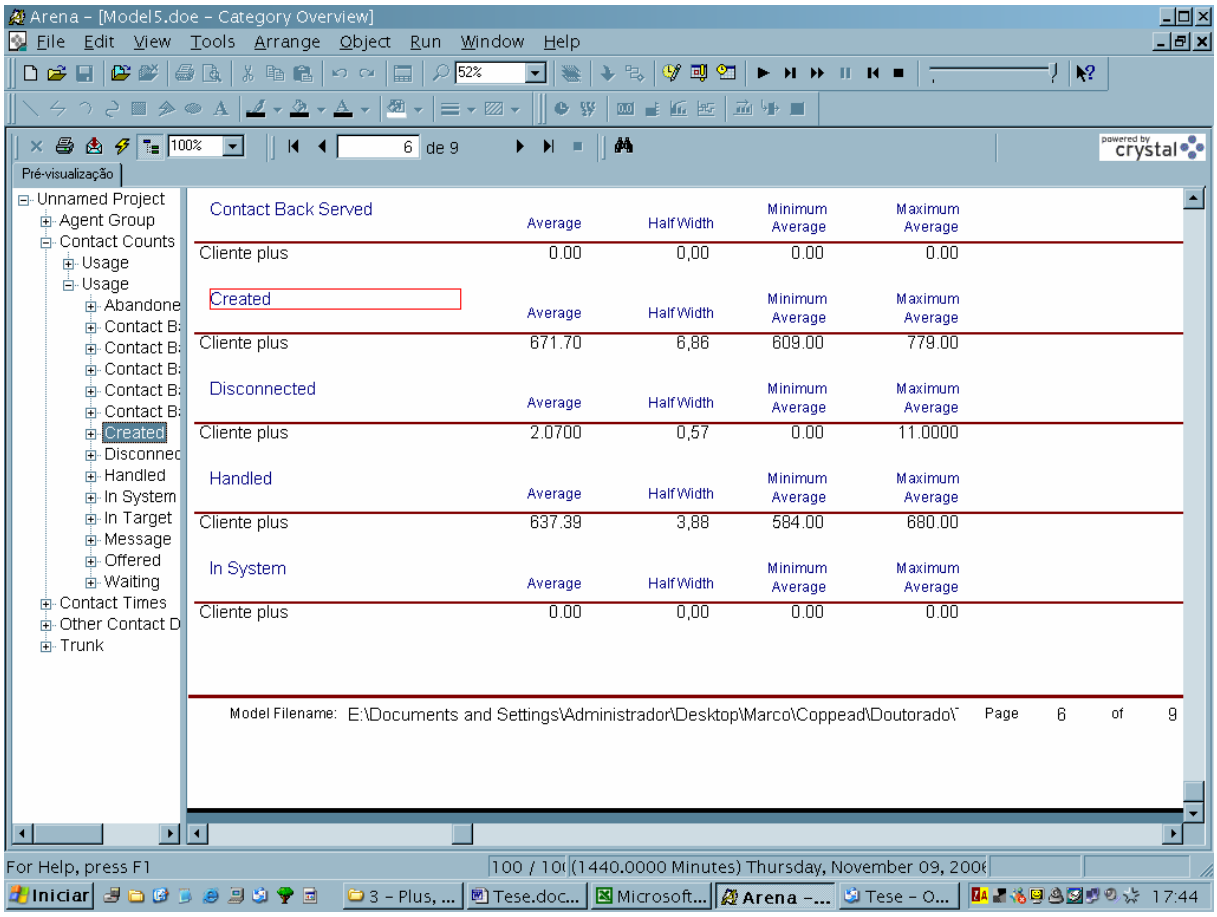
Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

A decisão a respeito de quantos atendentes serão designados e de que forma eles serão segmentados ou agregados em cada horário cabe totalmente à gerência de planejamento da Contax. Mas, pode ser que a empresa tenha interesse em estudar o impacto de outras variáveis que não estão sob seu controle (parâmetros) no nível de serviço e na taxa de abandono da central, o que é possível através da simulação, conforme indica a seção 3.5.

A análise de cenários pode ser utilizada para descobrir o que aconteceria com esses indicadores de performance se, por exemplo, o volume de chamadas em determinado horário fosse 10% maior do que o previsto.

As figuras 55 e 56 a seguir apresentam os indicadores para o cálculo do nível de serviço em um cenário com acréscimo no volume de ligações, mantendo o mesmo efetivo de atendimento.

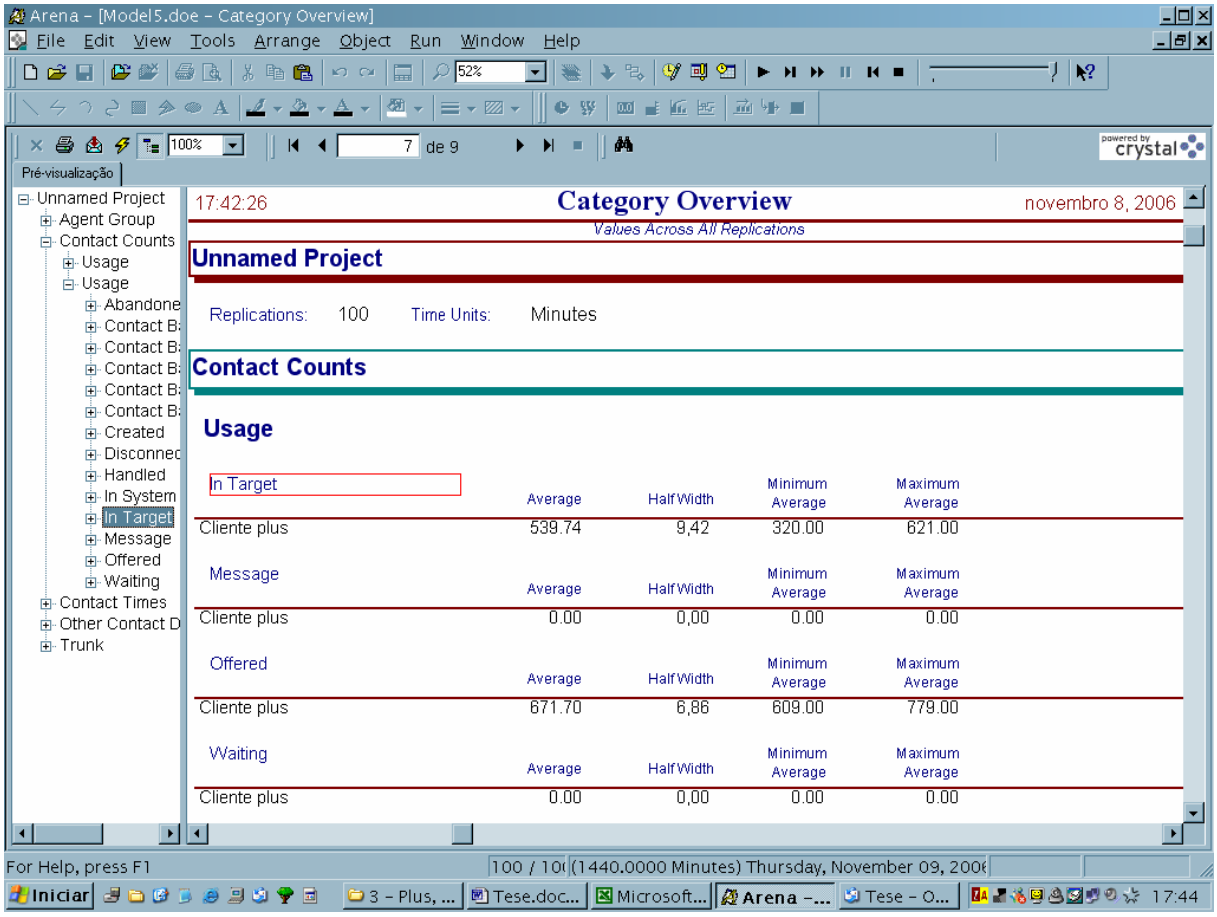
Figura 55 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% de demanda, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Neste cenário, 637 chamadas foram atendidas, em média; das quais, 540 – em média – em até 10 segundos. O nível de serviço para este cenário foi de 84,68%, conseqüentemente. Esse valor é ligeiramente menor que a meta de 85% estabelecida para os clientes *plus* e razoavelmente menor que os 93,31%, que teriam sido obtidos se a demanda se comportasse conforme o previsto.

Figura 56 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% de demanda, 100 replicações



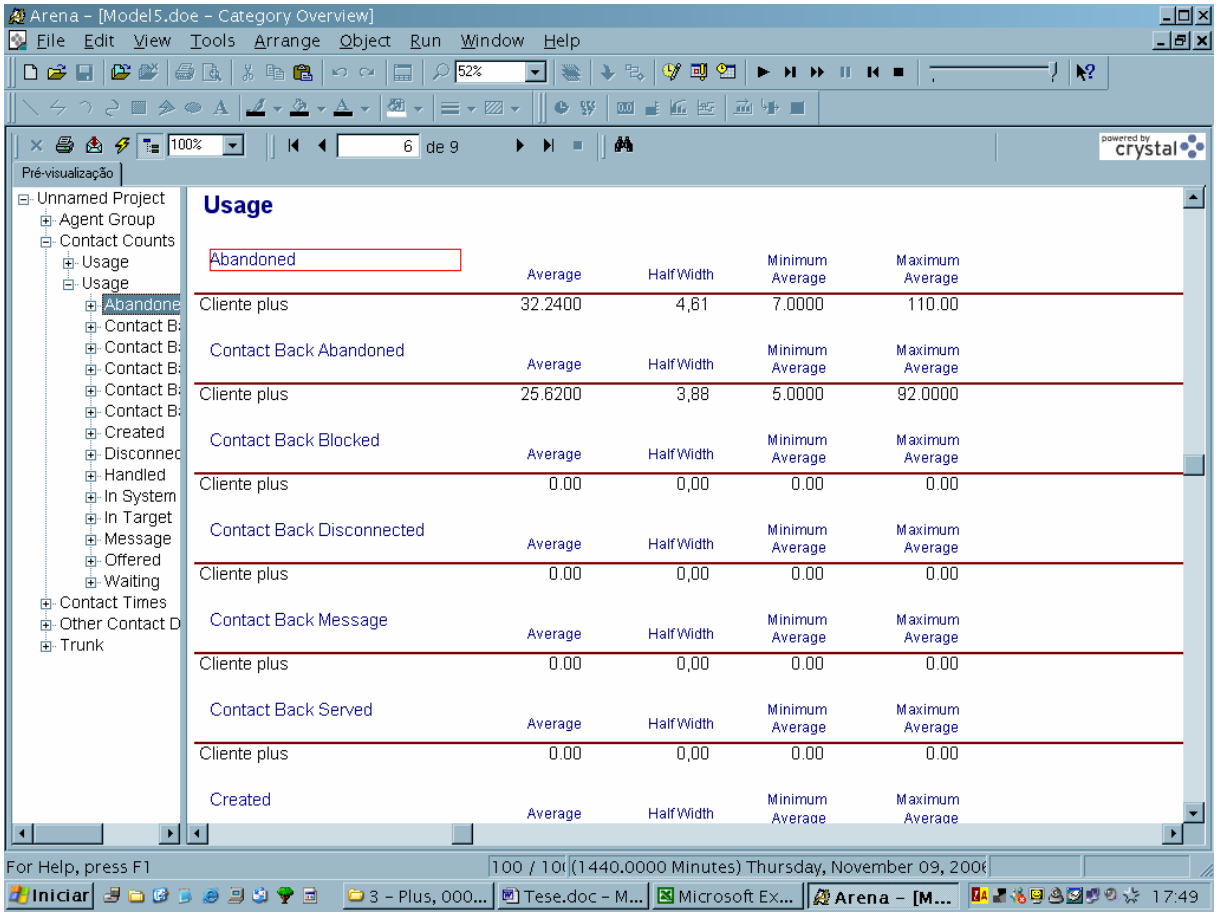
Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

O impacto deste acréscimo de demanda na taxa de abandono pode ser entendido a partir de uma informação extraída da figura 57 a seguir.

Das 672 chamadas geradas em cada replicação, 32,2 foram abandonadas, em média, pelos clientes, implicando em uma taxa de abandono igual a 4,80%, o que revela um grande impacto neste indicador de performance.

Assim, a simulação deste cenário mostrou que o efetivo original de atendentes (12) – em face de um imprevisível aumento de 10% na demanda – seria capaz de praticamente garantir a meta de nível de serviço (85%), mas comprometeria demais a taxa de abandono.

Figura 57 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% de demanda, 100 replicações

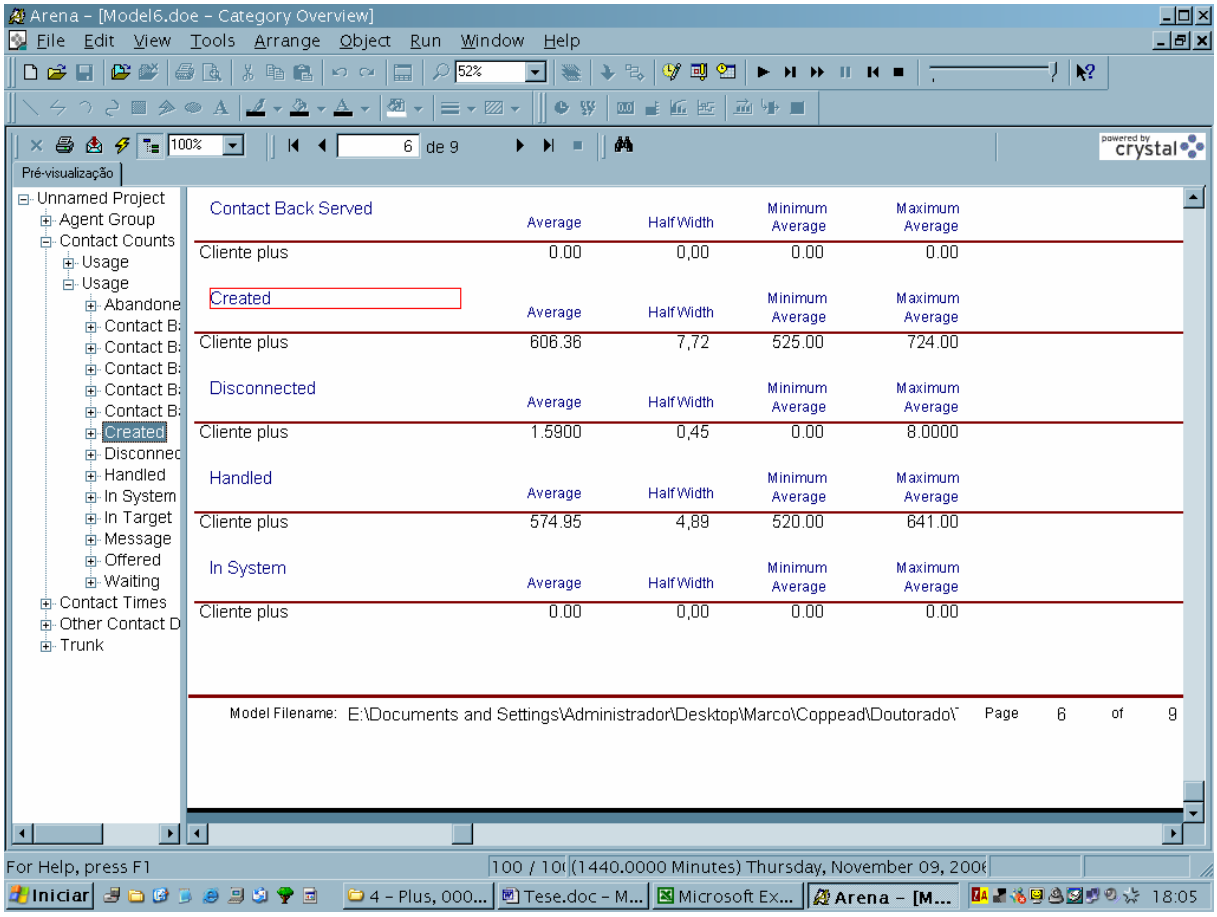


Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

E o que aconteceria com o nível de serviço e a taxa de abandono se a demanda tivesse sido subestimada (também em 10%), mas não em relação ao seu volume, e sim em relação ao TMA?

As figuras 58 e 59 a seguir mostram os valores necessários para a realização do cálculo do nível de serviço neste cenário sugerido.

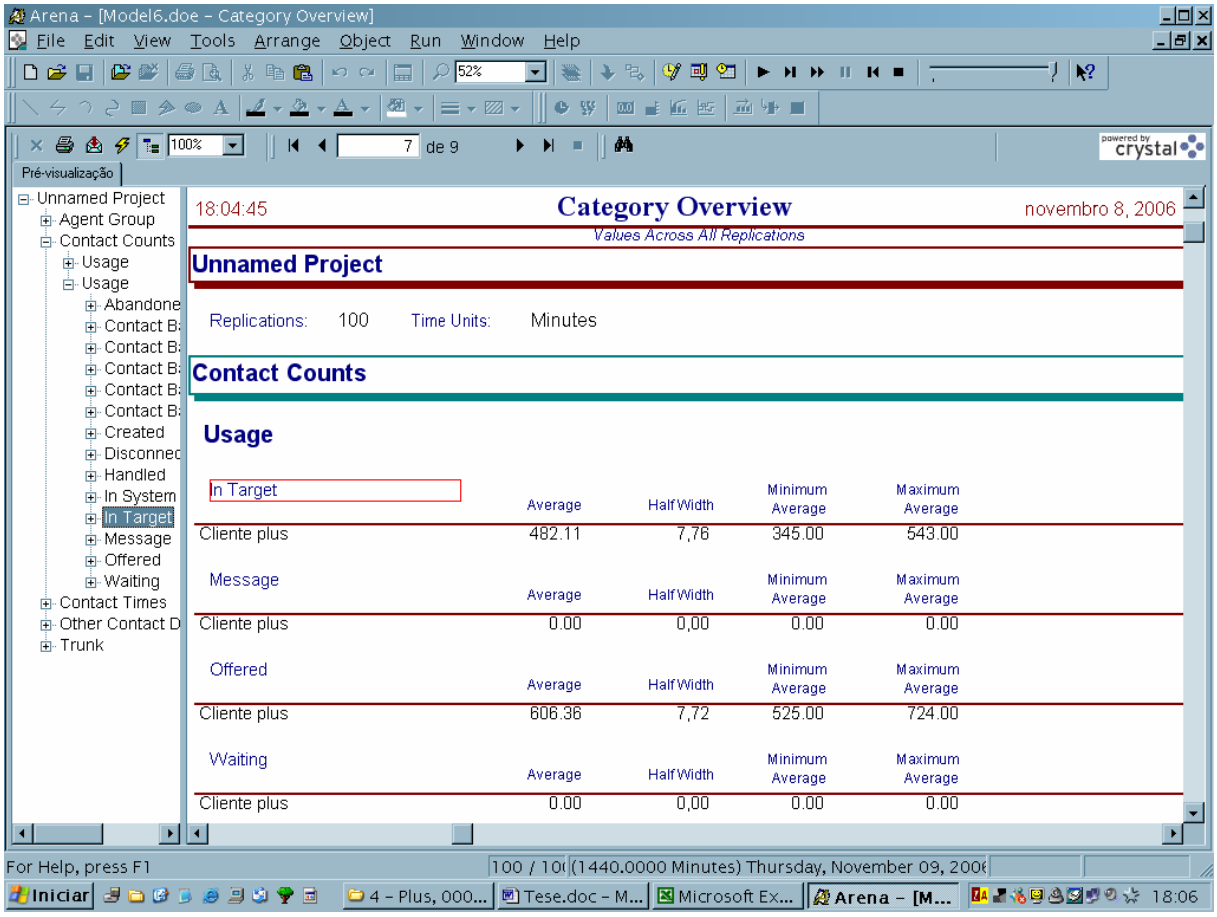
Figura 58 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% no TMA, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

No cenário em questão, 575 chamadas foram, em média, atendidas; destas, 482 (em média) em até 10 segundos. O nível de serviço resultante foi de 83,85%, um valor um pouco menor que a meta (85%) e razoavelmente menor que os 93,31%, que teriam sido obtidos se a demanda se comportasse conforme o previsto.

Figura 59 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% no TMA, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

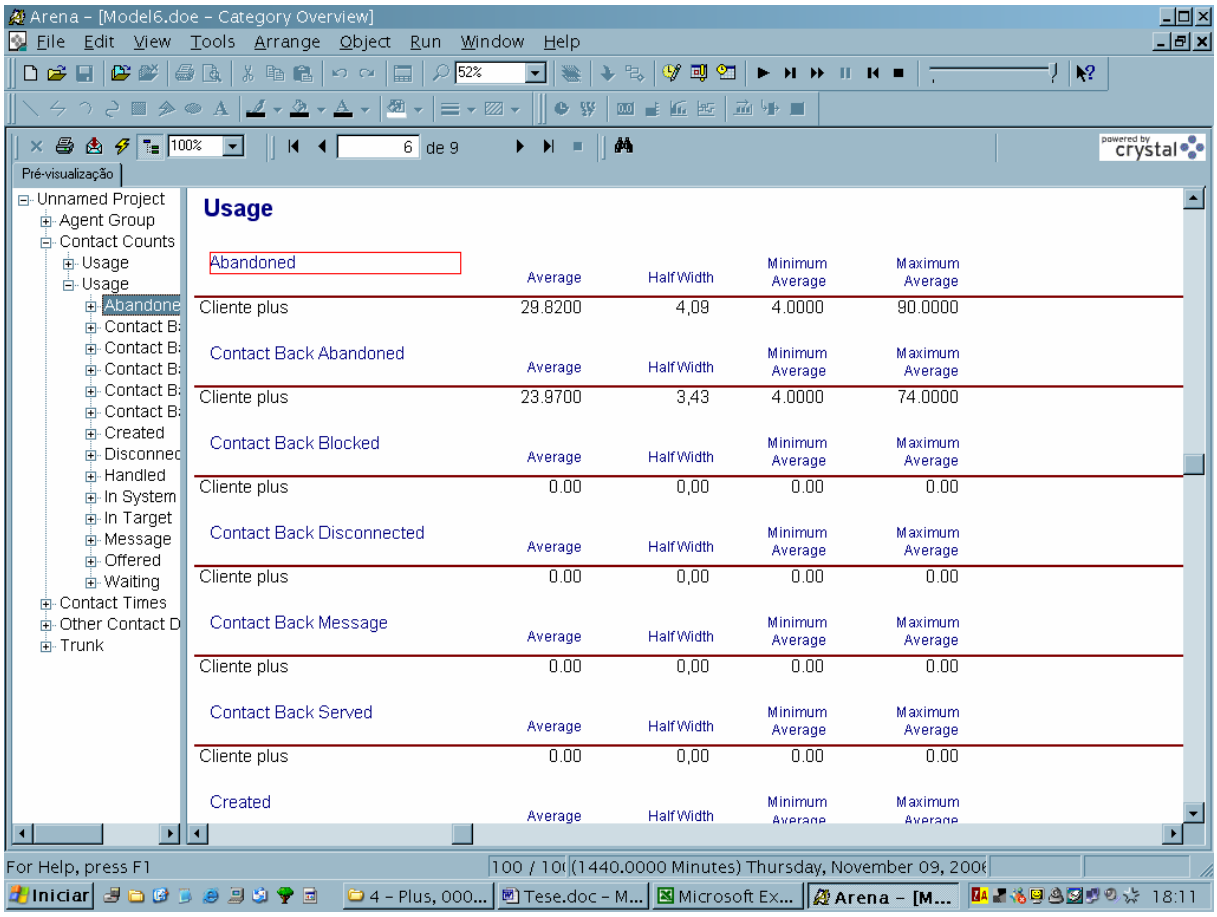
Para medir o impacto do aumento do TMA na taxa de abandono, é preciso recuperar uma informação exibida na figura 60 a seguir.

Das 606 chamadas geradas em cada replicação, 29,8 foram abandonadas pelos clientes, em média. A conseqüência é uma taxa de abandono igual a 4,92%, revelando um grande impacto nesta grandeza.

Como no cenário de aumento da quantidade de ligações, a simulação desta situação mostrou que o efetivo original de atendentes (12) – em face de um repentino aumento de 10% no tempo médio de atendimento – seria capaz de assegurar um nível de serviço quase igual à

meta de 85% (neste caso, um pouco mais distante), mas também incrementando (até um pouco mais) a taxa de abandono.

Figura 60 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com acréscimo de 10% no TMA, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

A partir da análise dos cenários com demanda mais intensa que a prevista, então, é possível concluir que variações não muito grandes (10%) em relação aos valores previstos podem impactar diretamente os indicadores de performance, especialmente a taxa de abandono. Esta conclusão faz exigir muito cuidado no momento de prever a quantidade de chamadas e o tempo médio de atendimento.

Outra revelação importante diz respeito ao fato – talvez surpreendente – do impacto ser ainda maior no caso de aumento do TMA, quando comparado ao impacto proporcionado por uma

diferença de mesma magnitude no volume de chamadas. Isso leva a crer que a acurácia na previsão do TMA é ainda mais importante do que a acurácia na previsão de volume.

Dessa forma, talvez seja interessante investigar o impacto de variações maiores do TMA (seja para mais ou para menos) no desempenho do sistema, através de uma análise de sensibilidade mais completa.

A simulação do modelo básico foi repetida no *software* algumas vezes, mas com diferentes valores para a média do tempo de atendimento (desde montantes 30% menores até montantes 30% maiores), de onde os resultados relevantes foram coletados. Eles foram, então, organizados na tabela 39 a seguir, que também calcula e apresenta os principais indicadores de desempenho (nível de serviço – percentual de ligações atendidas em até 10 segundos – e taxa de abandono) de cada cenário.

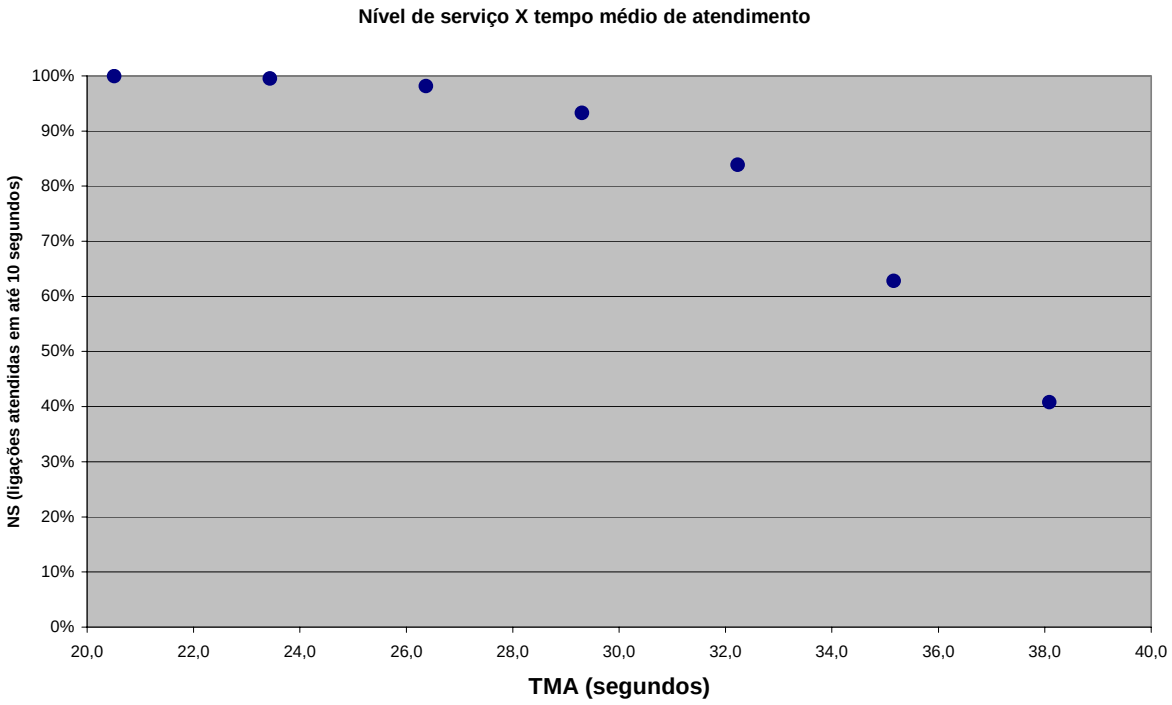
Tabela 39 – Nível de serviço e taxa de abandono para diferentes valores do tempo médio de atendimento (TMA), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações

TMA (seg)	Variação	Ligações				Nível de serviço	Taxa de abandono
		geradas	atendidas	em até 10 seg	abandonadas		
20,5	-30%	583	582	582	0,84	99,92%	0,14%
23,4	-20%	586	583	580	2,68	99,50%	0,46%
26,4	-10%	589	582	571	6,34	98,13%	1,08%
29,3	**-**	**595**	**579**	**541**	**14,52**	**93,31%**	**2,44%**
32,2	+10%	606	575	482	29,82	83,85%	4,92%
35,2	+20%	640	567	356	68,39	62,80%	10,69%
38,1	+30%	683	548	223	128,01	40,77%	18,74%

Fonte: Tabela elaborada pelo pesquisador a partir de resultados obtidos pelo *software* Arena Contact Center

À medida que o TMA aumenta, os clientes demoram mais (em média) sendo atendidos, as filas vão crescendo e o nível de serviço vai despencando, atingindo um patamar inaceitável quando o TMA é 30% maior do que o valor básico. Essa tendência evidenciada na tabela 39 anterior pode ser mais bem visualizada através do gráfico apresentado na figura 61 a seguir.

Figura 61 – Nível de serviço para diferentes valores do tempo médio de atendimento (TMA), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações

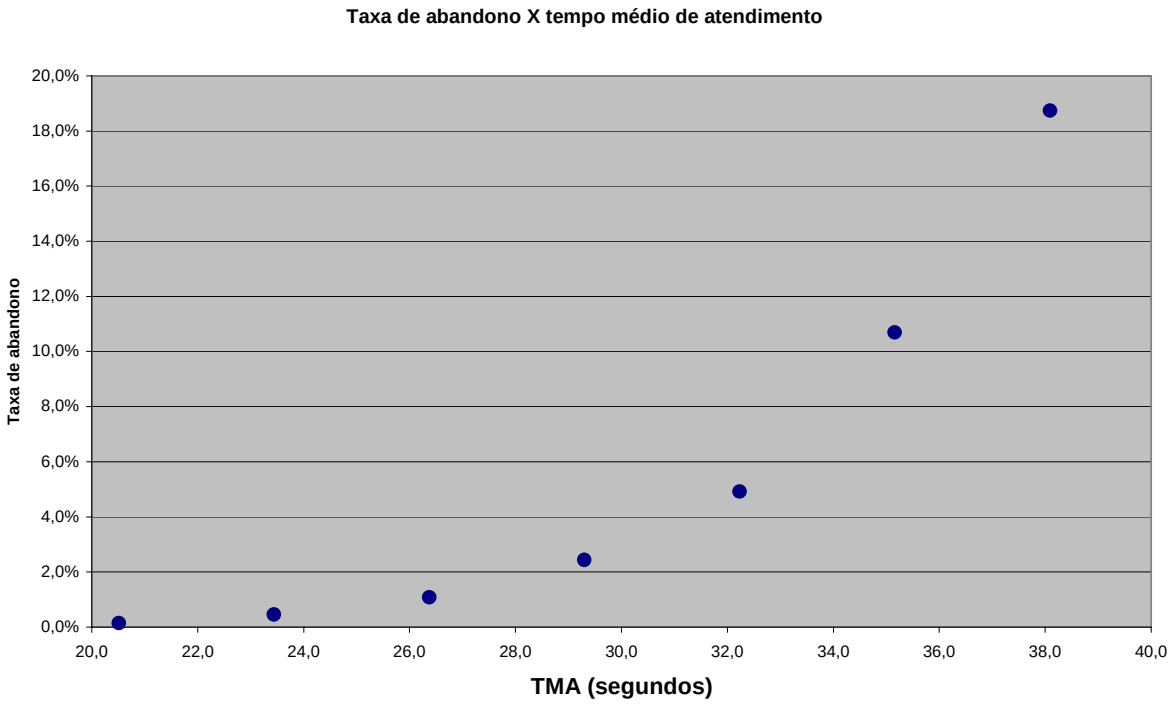


Fonte: Gráfico elaborado pelo pesquisador

Pelo mesmo motivo, mais clientes resolvem não mais esperar na fila, fazendo a taxa de abandono subir rapidamente, chegando a ponto de praticamente 1 a cada 5 clientes abandonarem a fila antes de serem atendidos, quando o valor do TMA está 30% acima do seu montante básico. Essa tendência pode ser observada no gráfico apresentado na figura 62 a seguir.

Portanto, valores do TMA mais altos que o previsto deterioram rapidamente o desempenho do sistema em termos do nível de serviço e da taxa de abandono (esta última, principalmente) revelando um enorme impacto potencial do aumento do TMA nos resultados mais importantes. Dessa maneira, a Contax deve se esforçar para evitar ao máximo que o tempo médio de atendimento cresça, conscientizando seus atendentes acerca das destrutivas conseqüências do aumento do valor desta grandeza.

Figura 62 – Taxa de abandono para diferentes valores do tempo médio de atendimento (TMA), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações



Fonte: Gráfico elaborado pelo pesquisador

Analisando a parte superior da tabela 39 anterior, é possível concluir que os indicadores de desempenho melhoram sensivelmente com uma redução de apenas 10% no TMA: o nível de serviço aumenta em quase 5 pontos percentuais, ultrapassando os 98% e a taxa de abandono cai para menos do que a metade (1,08%). Isso sugere que pode valer a pena investir em treinamento para os agentes, de forma a tentar reduzir um pouco o tempo de atendimento.

Infelizmente, também é verdadeiro que reduções ainda maiores no TMA não encerram vantagens tão marcantes (especialmente para o nível de serviço, que já se encontraria em um nível próximo do ótimo) para o sistema. Em outras palavras, é possível que o custo envolvido para diminuir o TMA em 10% seja compensando pelos benefícios resultantes desta redução, mas é difícil crer que o mesmo ocorra para reduções mais drásticas desta grandeza.

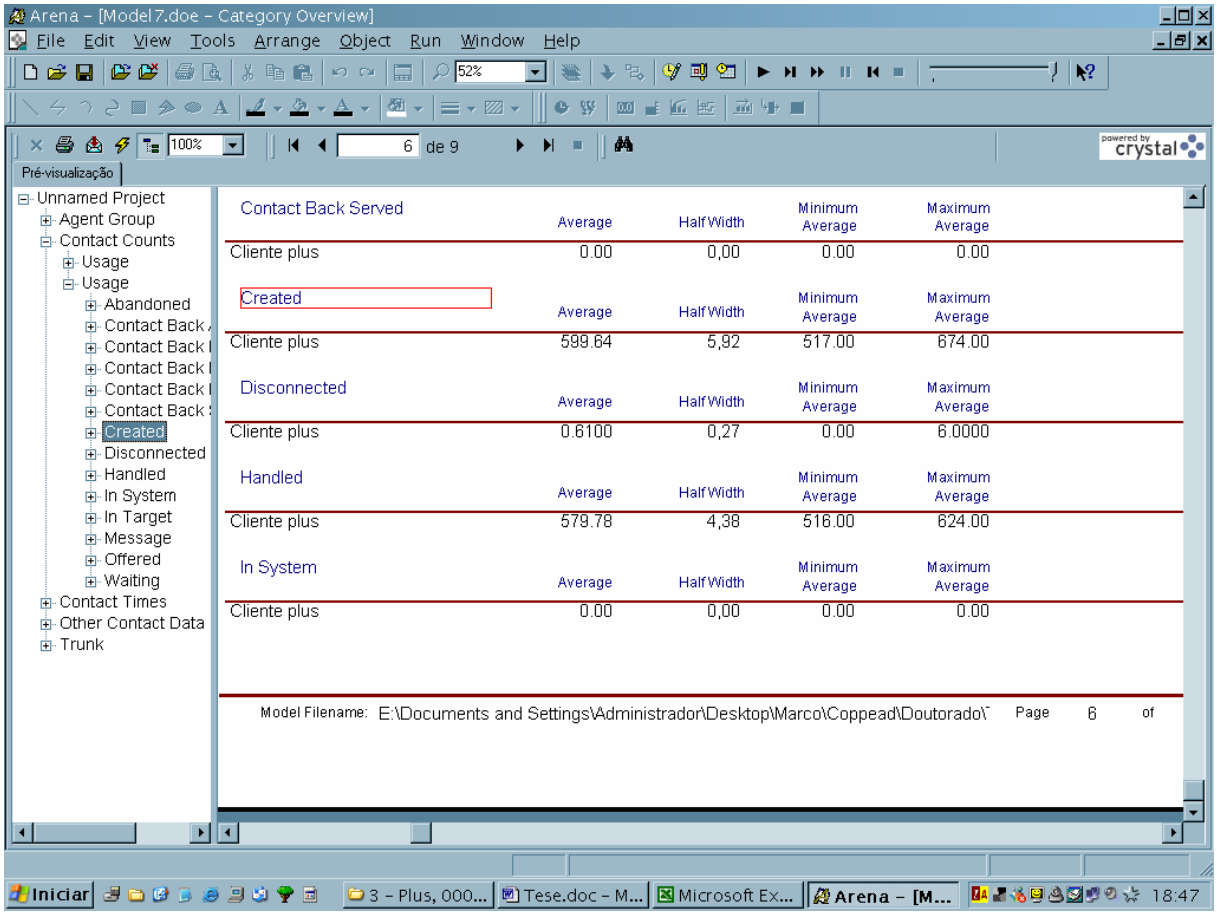
De forma semelhante, cabem outras indagações do mesmo tipo:

- o que aconteceria com a taxa de abandono se o cliente ficasse mais impaciente e passasse a abandonar as chamadas caso demorasse mais de, por exemplo, 1,5 minutos (ao invés de 2,5), em média, para ser atendido?

- qual o impacto desta mudança no nível de serviço?

As figuras 63 e 64 a seguir mostram os números necessários para o cálculo do nível de serviço neste novo cenário proposto.

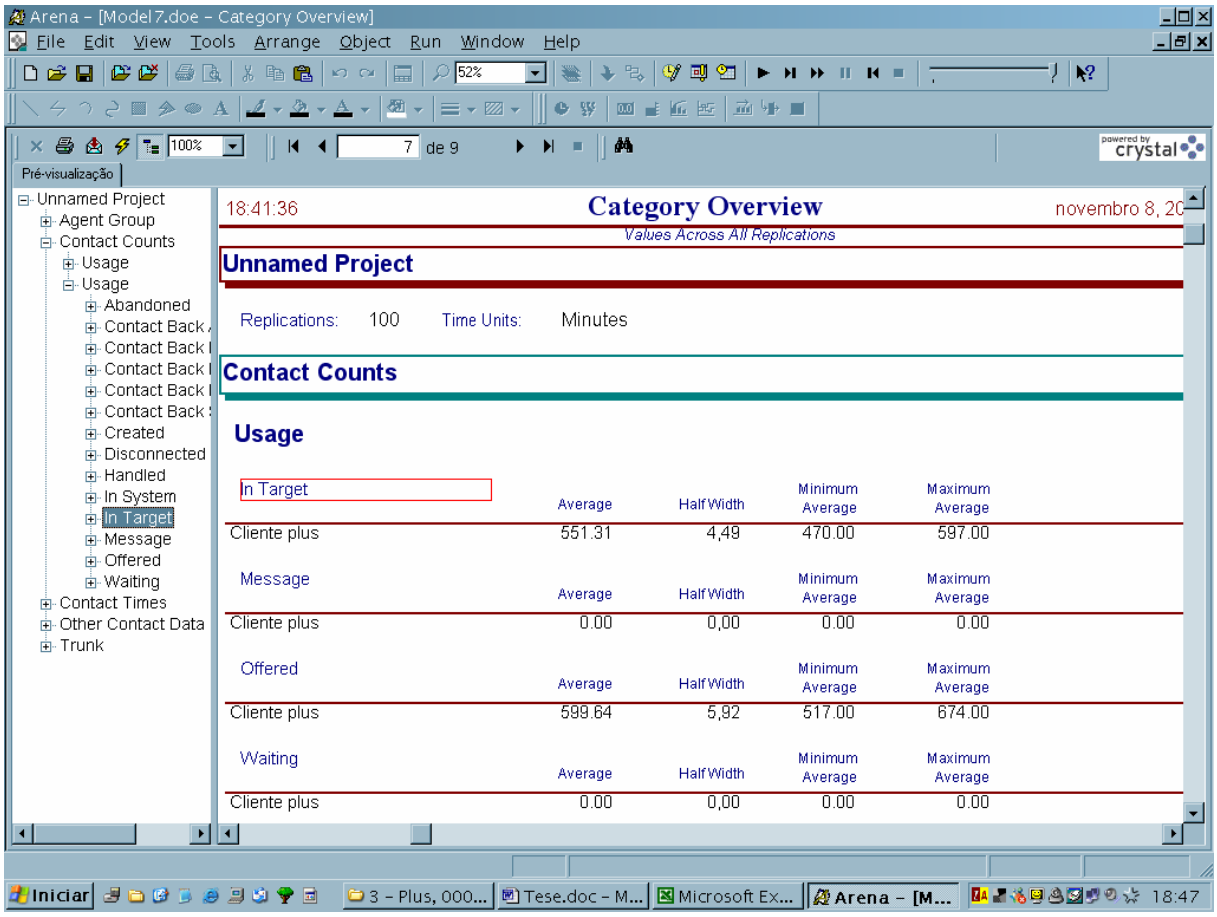
Figura 63 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com tempo médio de espera antes do abandono = 1,5 minutos, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Nele, 580 chamadas foram atendidas, em média; das quais, 551 – em média – em até 10 segundos. O nível de serviço para este cenário foi de 95,09%, conseqüentemente. Esse valor é um pouco maior que os 93,31% (que teriam sido obtidos com o comportamento anterior de abandono por parte dos clientes), e também razoavelmente superior à meta de 85%.

Figura 64 – Chamadas atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com tempo médio de espera antes do abandono = 1,5 minutos, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Mesmo aumentando em 40% "a impaciência do cliente" – ou seja, diminuindo o tempo médio de espera antes do abandono de 2,5 minutos para 1,5 minutos – o impacto no nível de serviço foi pequeno. Talvez fosse esperado um aumento maior nesta grandeza, visto que, com mais clientes abandonando a fila, seria natural que os remanescentes fossem atendidos de forma mais rápida.

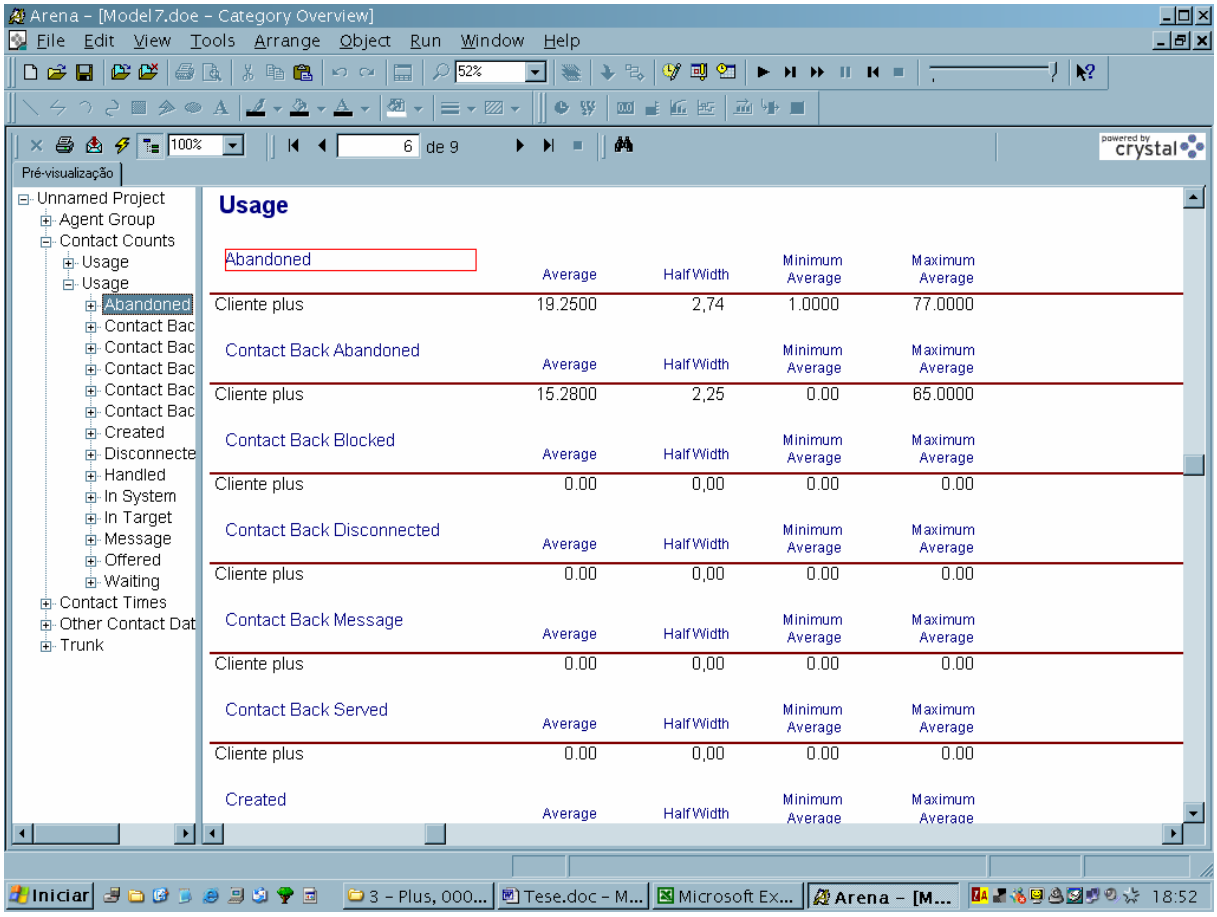
O que sucede é que, de acordo com o modelo, 80% das ligações abandonadas retornam à fila alguns minutos depois, congestionando novamente o sistema e não deixando o nível de serviço aumentar tanto. Esse tipo de análise e conclusão seriam praticamente impossíveis através de abordagens analíticas, que ignoram o comportamento de abandono (de acordo com o que foi mencionado na seção 3.5), só sendo viáveis através de métodos experimentais, como a simulação.

Se o interesse da gerência residir apenas no nível de serviço, deixa até de ser interessante empreender um esforço muito grande para prever com perfeita exatidão o tempo que os clientes costumam esperar, em média, antes de abandonar a chamada. Isto porque foi verificado que uma mudança nada pequena (40%) nesse tempo médio é incapaz de impactar intensamente o nível de serviço.

Se, no entanto, houver interesse também em monitorar a taxa de abandono, é fundamental estudar o impacto do novo cenário nesta grandeza, a partir da análise de uma informação a ser extraída da figura 65 a seguir.

Das 600 chamadas geradas em cada replicação, 19,3 foram abandonadas, em média, pelos clientes, implicando em uma taxa de abandono igual a 3,21%, um pouco maior do que os 2,85% obtidos anteriormente.

Figura 65 – Chamadas abandonadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com tempo médio de espera antes do abandono = 1,5 minutos, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Para verificar a sensibilidade dos indicadores de performance em relação à variável tempo médio de espera antes do abandono de forma mais completa, a mesma simulação foi repetida algumas vezes no *software*, com diferentes valores para esta variável. Foram coletados os *outputs* necessários para o cálculo dos principais indicadores de desempenho (nível de serviço – percentual de ligações atendidas em até 10 segundos – e taxa de abandono), que estão apresentados na tabela 40 a seguir.

Como pode ser observado, a taxa de abandono é razoavelmente sensível em relação ao tempo médio de espera, principalmente nos patamares mais baixos desta variável. Sendo assim, a sua correta consideração mostra-se importante para a obtenção de resultados acurados, não

obstante o fato do nível de serviço revelar pouca sensibilidade a mudanças na mesma variável (fato ocorrido provavelmente pelo motivo cogitado anteriormente).

Tabela 40 – Nível de serviço e taxa de abandono para diferentes valores do tempo médio de espera (antes do abandono), horário das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, 100 replicações

Tempo médio de espera (minutos)	Ligações				Nível de serviço	Taxa de abandono
	geradas	atendidas	em até 10 seg	abandonadas		
0,5	624	578	562	46,65	97,26%	7,47%
1,5	600	580	551	19,25	95,09%	3,21%
2,5	**595**	**579**	**541**	**14,52**	**93,31%**	**2,44%**
3,5	593	582	540	10,02	92,75%	1,69%
4,5	591	582	534	9,10	91,91%	1,54%

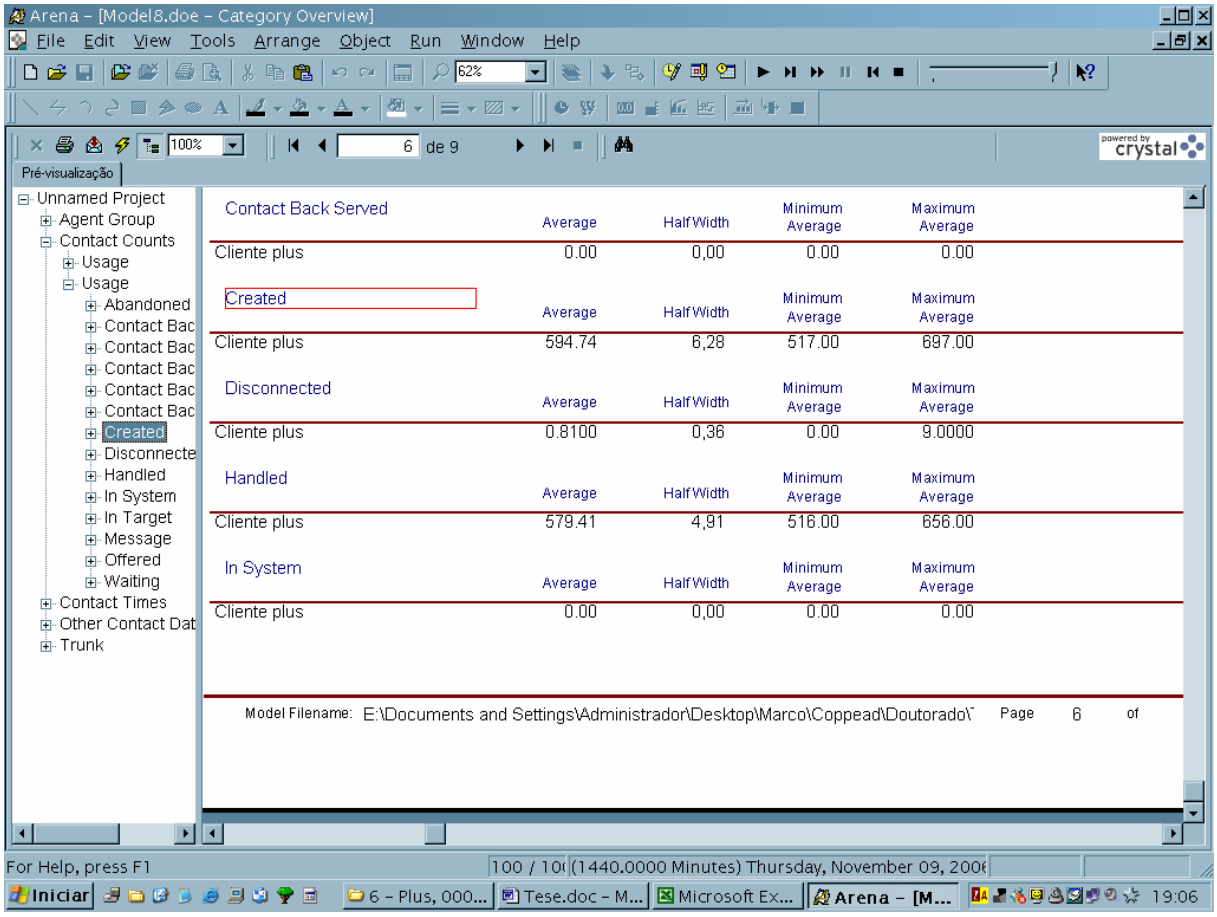
Fonte: Tabela elaborada pelo pesquisador a partir de resultados obtidos pelo *software* Arena Contact Center

O gerente de planejamento também pode estar interessado em saber o que aconteceria se a empresa contratante ficasse mais exigente em relação ao nível de serviço e redefinisse seu conceito, passando este a corresponder ao percentual de clientes atendidos em até 5 segundos (ao invés de 10 segundos).

As figuras 66 e 67 a seguir revelam os valores para o cálculo do novo nível de serviço, após essa redefinição no seu conceito.

Neste cenário, 579 chamadas foram atendidas, em média; das quais, 489 – em média – em até 5 segundos. O nível de serviço para este cenário foi de 84,44%, por conseguinte. Esse valor é razoavelmente menor que os 93,31% obtidos com a definição original de nível de serviço e, o que é mais importante, um pouco menor que a meta de 85% estabelecida para os clientes *plus*.

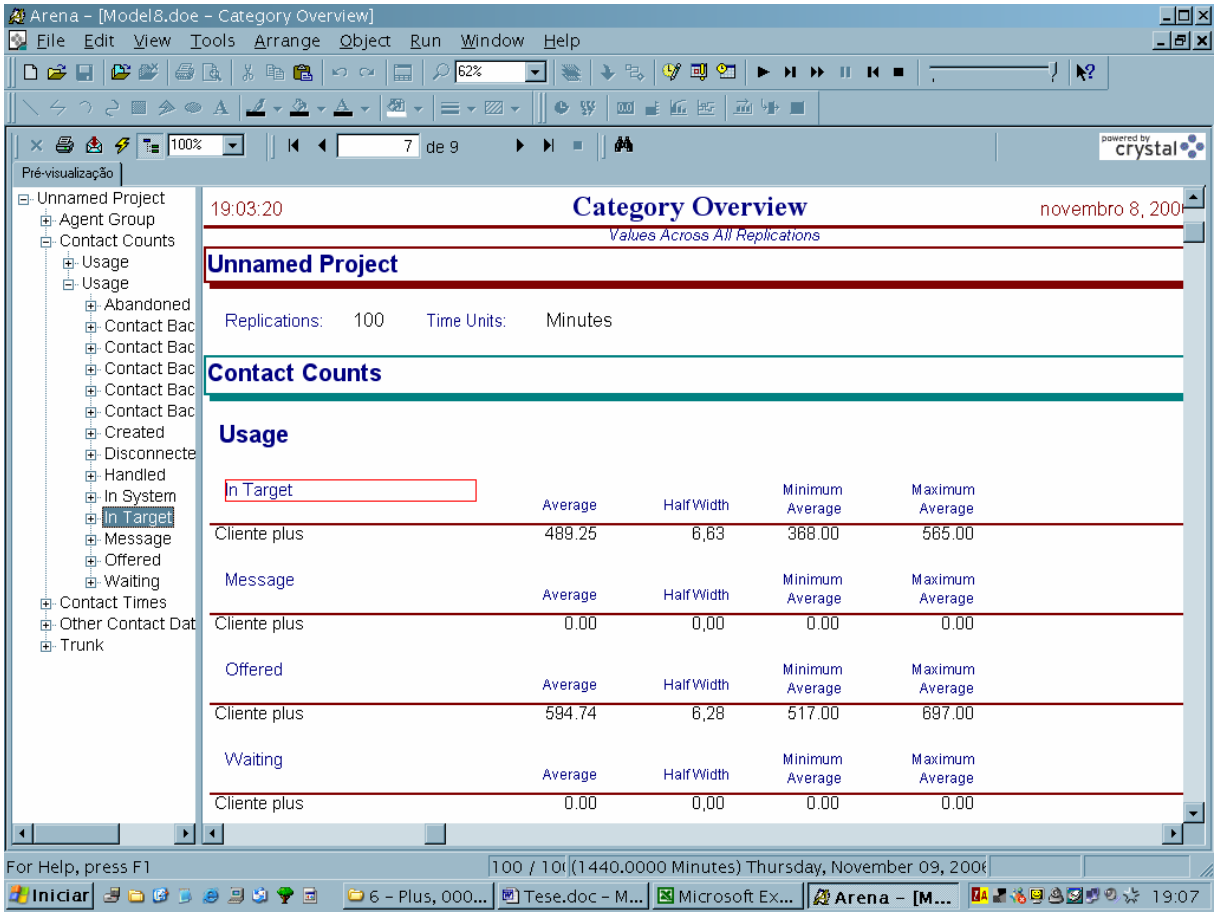
Figura 66 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com NS = % de clientes atendidos em até 5 segundos, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Essa comparação revela que uma redefinição no conceito de nível de serviço pode impactar de forma não muito branda esse indicador de desempenho, o que era razoavelmente previsível. É possível até que – exatamente como aconteceu no exemplo ilustrado – a atual configuração de atendentes não seja mais suficiente para proporcionar um nível de serviço condizente com a meta pré-estabelecida. Neste caso, pode ser importante descobrir quantos agentes adicionais serão necessários para que este indicador de performance retorne a patamares acima da meta.

Figura 67 – Chamadas atendidas em até 5 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com NS = % de clientes atendidos em até 5 segundos, 100 replicações

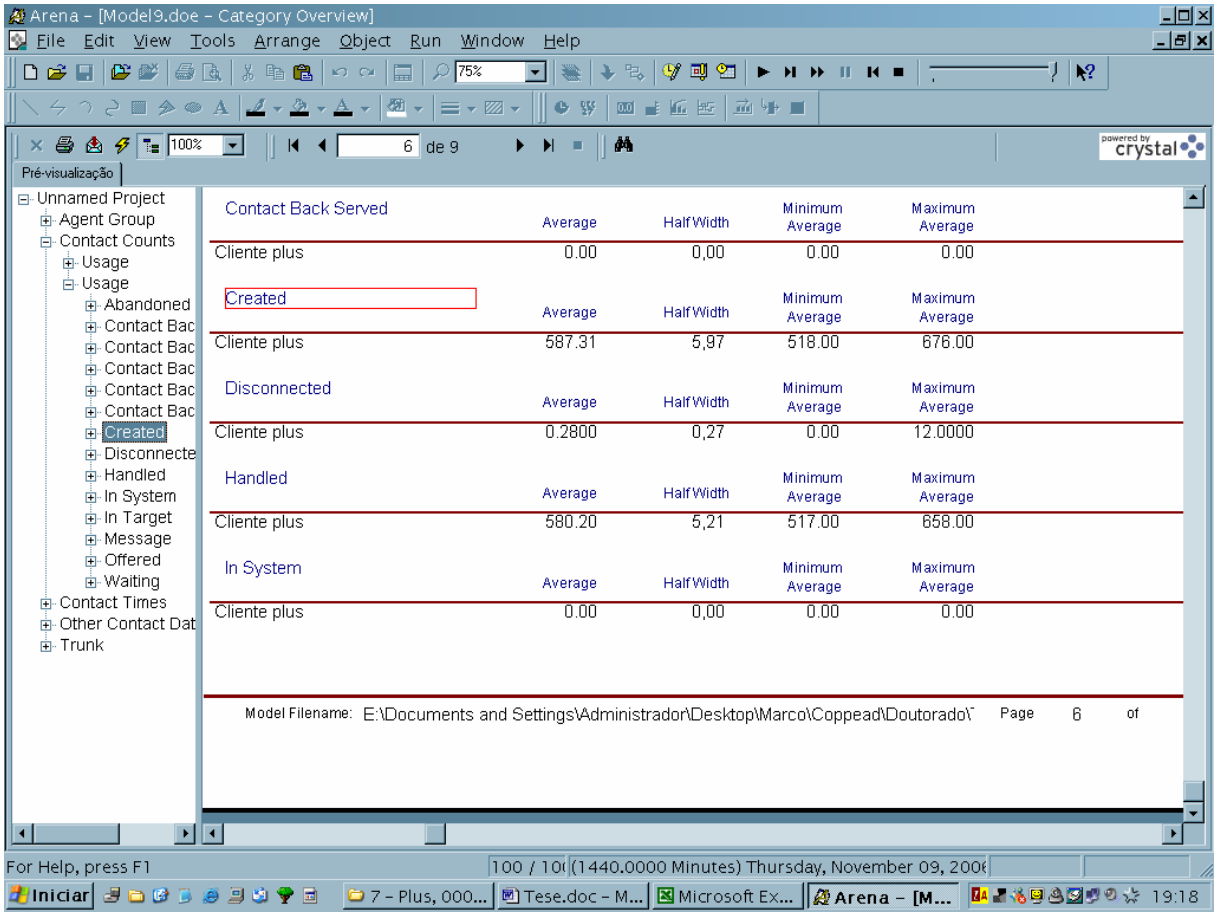


Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Para tal, primeiro é necessário verificar se a adição de 1 agente será suficiente para o objetivo ser atingido. As figuras 68 e 69 a seguir apresentam os indicadores que entram no cálculo do nível de serviço (segundo o seu novo conceito), mas em um cenário com 13 atendentes efetivos.

Foram atendidas 580 chamadas – em média – neste cenário, sendo que 534 (em média) delas em até 5 segundos. O nível de serviço resultante foi de 91,99%, um valor acima da meta (85%) e dos 84,44% obtidos com 12 agentes, mas um pouco menor que os 93,31% originais.

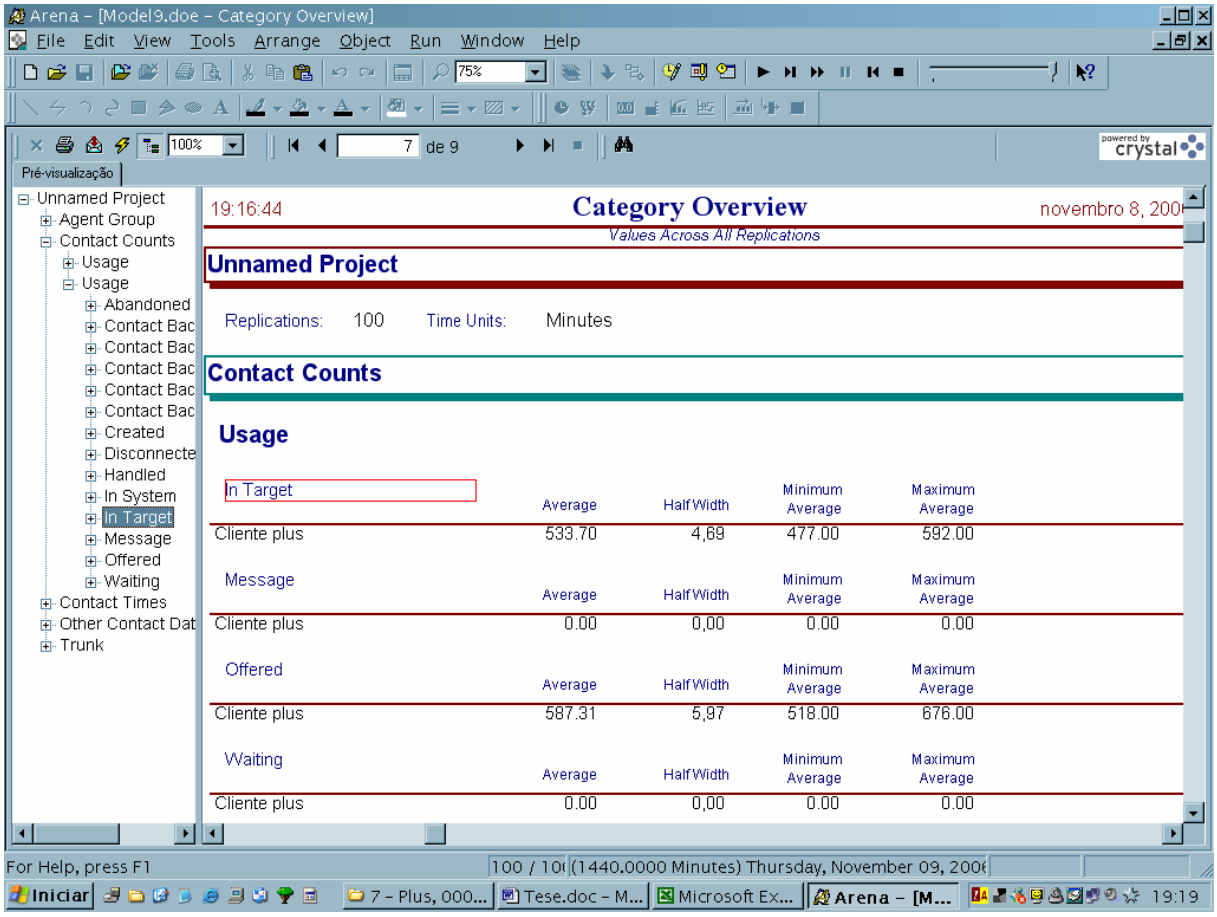
Figura 68 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 13 atendentes e NS = % de clientes atendidos em até 5 segundos, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Isso mostra que a alocação de um agente adicional é capaz de fazer o nível de serviço voltar a atingir a meta, mas seus benefícios não chegam a compensar o impacto negativo para este indicador de desempenho causado pela reformulação do seu conceito.

Figura 69 – Chamadas atendidas em até 5 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), cenário com 13 atendentes e NS = % de clientes atendidos em até 5 segundos, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

E o que aconteceria se a central dispusesse de uma quantidade limitada de linhas disponíveis no seu tronco? Na verdade, isso sempre ocorre, mas muitas vezes essa quantidade é tão grande, que acaba não se constituindo em uma restrição. Mas e se essa quantidade for apenas um pouco maior do que a quantidade de agentes dimensionados?

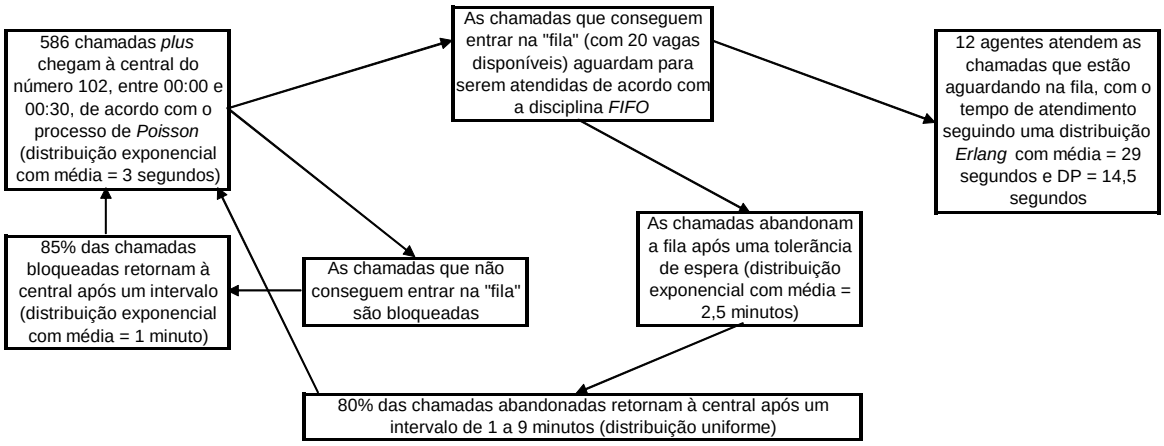
Em outras palavras, o que ocorreria se o sistema tivesse uma capacidade limitada e baixa? Se houvesse, por exemplo, apenas 20 linhas disponíveis, não existindo espaço no sistema, portanto, para mais clientes (sendo atendidos ou aguardando na fila) do que essa quantidade?

Certamente alguns clientes seriam bloqueados antes de entrar no sistema, isto é, impedidos até mesmo de aguardar na fila (receberiam um sinal de ocupado ou teriam sua ligação interrompida). Mas quantos?

E se uma parcela destes clientes retornasse a ligação poucos instantes após este bloqueio? E qual o impacto dessa limitação nos indicadores de performance? As respostas a essas perguntas muito dificilmente poderiam ser obtidas através de modelos analíticos, mas podem ser conseguidas com certa facilidade através da simulação.

O modelo para simular este cenário levou em conta uma quantidade limitada (20) de linhas disponíveis, assumindo a premissa de que 85% dos clientes bloqueados voltariam a tentar entrar em contato pouco tempo depois (este tempo foi modelado através de uma distribuição exponencial com média de 1 minuto). Ele está esquematizado na figura 70 a seguir.

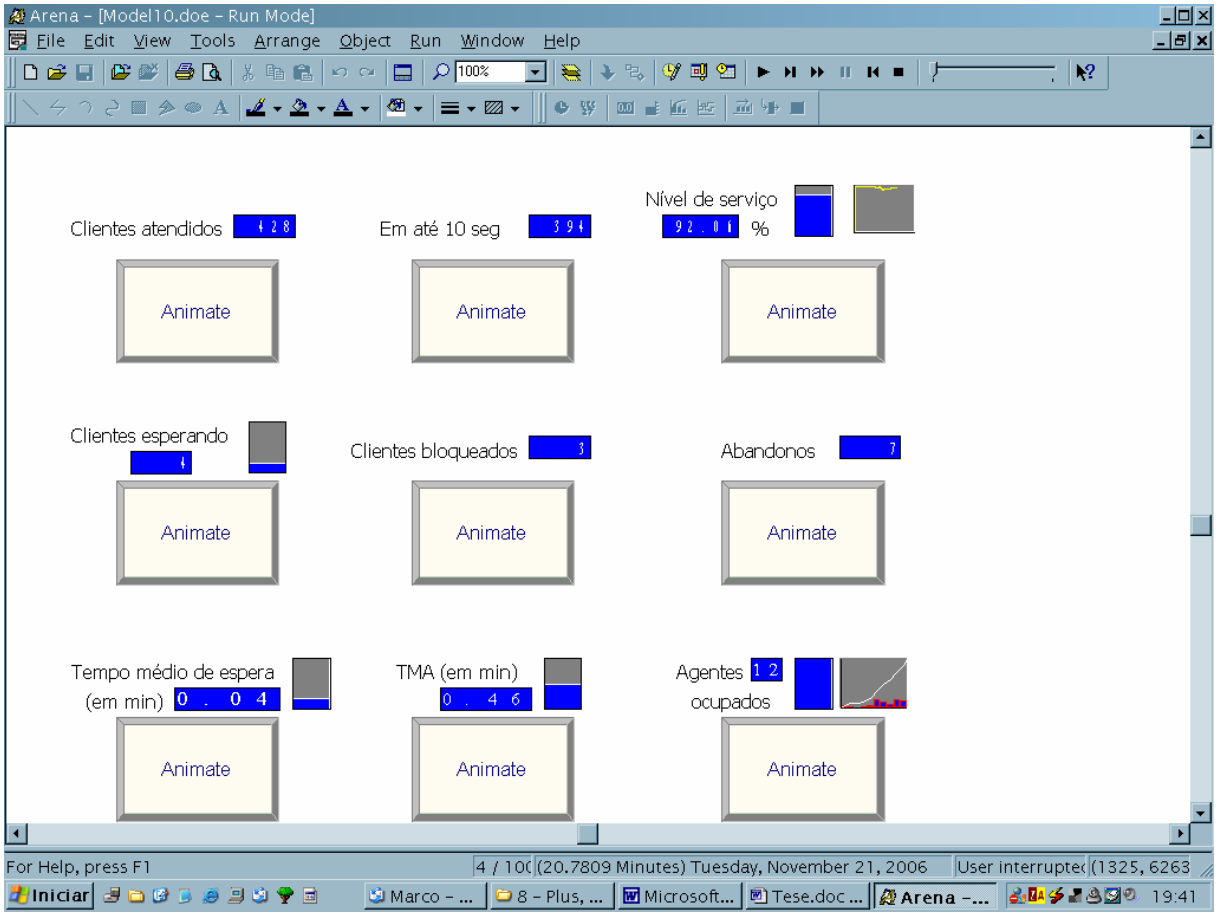
Figura 70 – Diagrama de processos do modelo de simulação com quantidade limitada (20) de linhas disponíveis, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes



Fonte: Diagrama elaborado pelo pesquisador

Este cenário mais completo (possibilitando diferentes *status* para as chamadas: atendidas, atendidas dentro do limite, aguardando, bloqueadas e abandonadas) foi escolhido para ilustrar a apresentação dinâmica dos indicadores de desempenho que o software de simulação é capaz de oferecer, conforme pode ser observado na figura 71 a seguir.

Figura 71 – Indicadores dinâmicos do modelo de simulação com quantidade limitada (20) de linhas disponíveis, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes

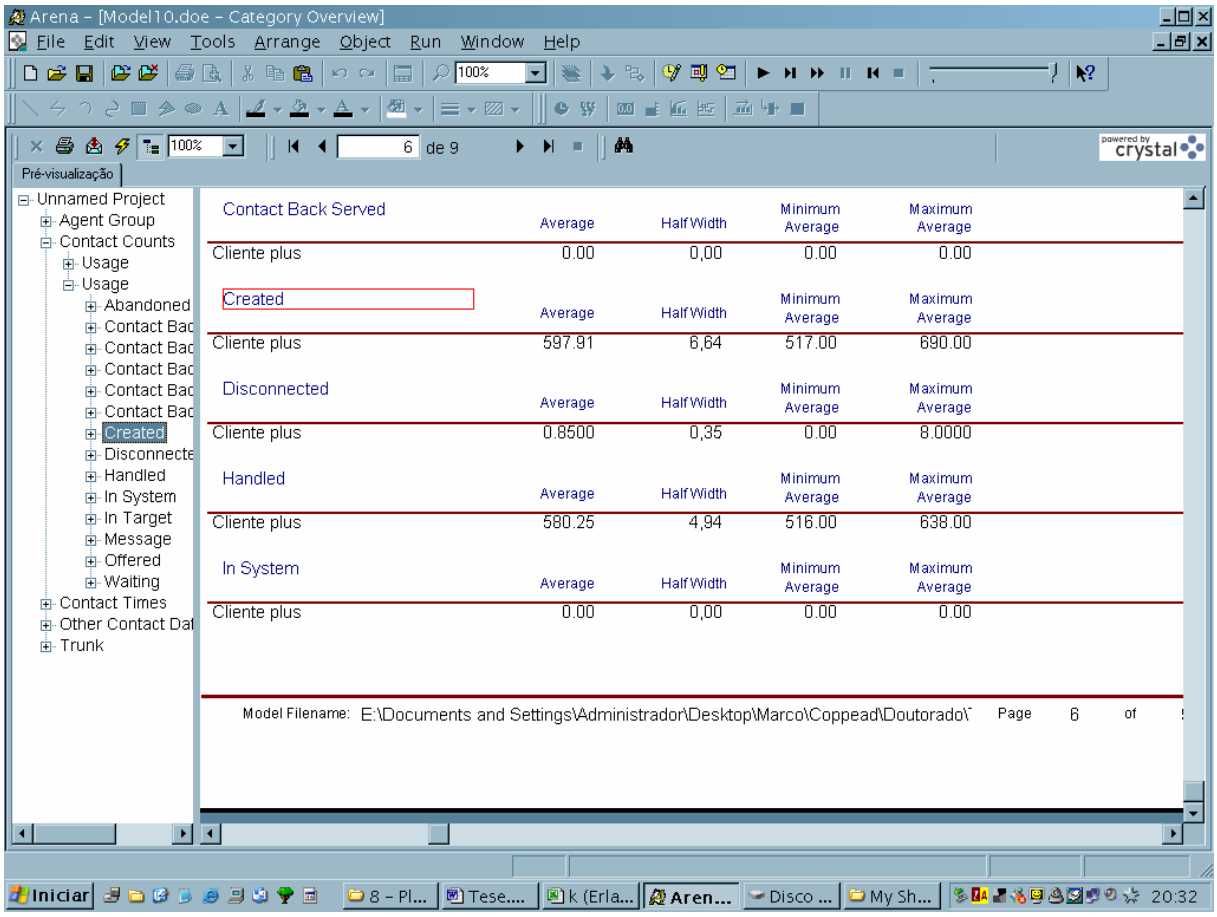


Fonte: Tela capturada durante a simulação do modelo pelo *software* Arena Contact Center

Como pode ser visto, diversos indicadores podem ser acompanhados – à medida que vão variando dinamicamente enquanto a simulação vai sendo executada – para este cenário: a quantidade de clientes atendidos, atendidos em até 10 segundos, o conseqüente nível de serviço (seu valor no momento, na forma numérica e gráfica – a barra de nível, além do seu comportamento ao longo da replicação), o número de clientes esperando na fila (na forma numérica e gráfica), de clientes bloqueados e que abandonaram a fila, os tempos médios de espera e de atendimento (na forma numérica e gráfica), e a quantidade de agentes ocupados (seu valor no momento, na forma numérica e gráfica, além de um histograma registrando seu comportamento ao longo da replicação).

No final das replicações, os relatórios sumarizam os resultados consolidados, estando os primeiros apresentados na figura 72 a seguir.

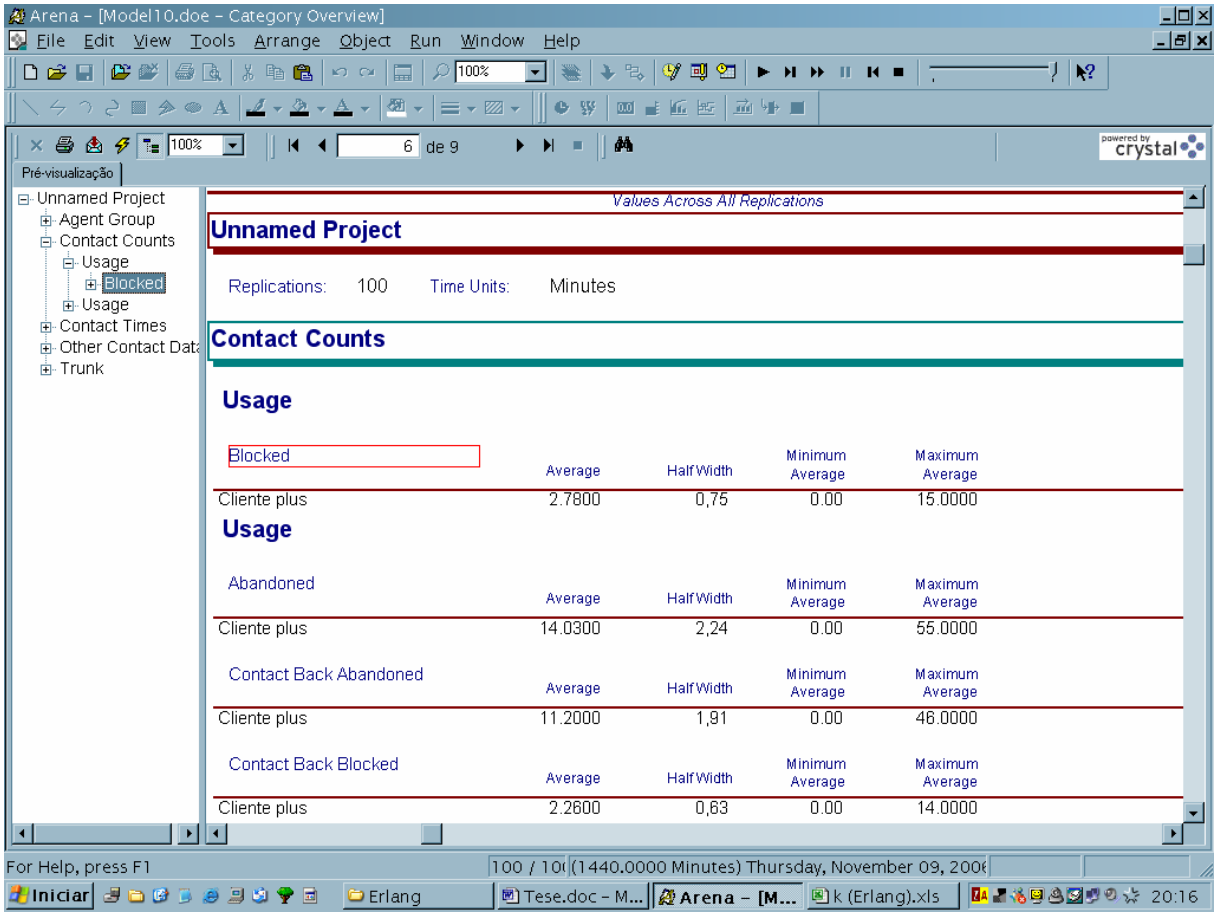
Figura 72 – Chamadas geradas e atendidas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Em média, 598 chamadas foram geradas em cada replicação; dessas, 580 – em média – foram efetivamente atendidas pelos agentes em cada replicação. Dentre o total de chamadas geradas, as quantidades que foram bloqueadas e abandonadas podem ser visualizadas na figura 73 a seguir.

Figura 73 – Chamadas bloqueadas, abandonadas e retornadas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações



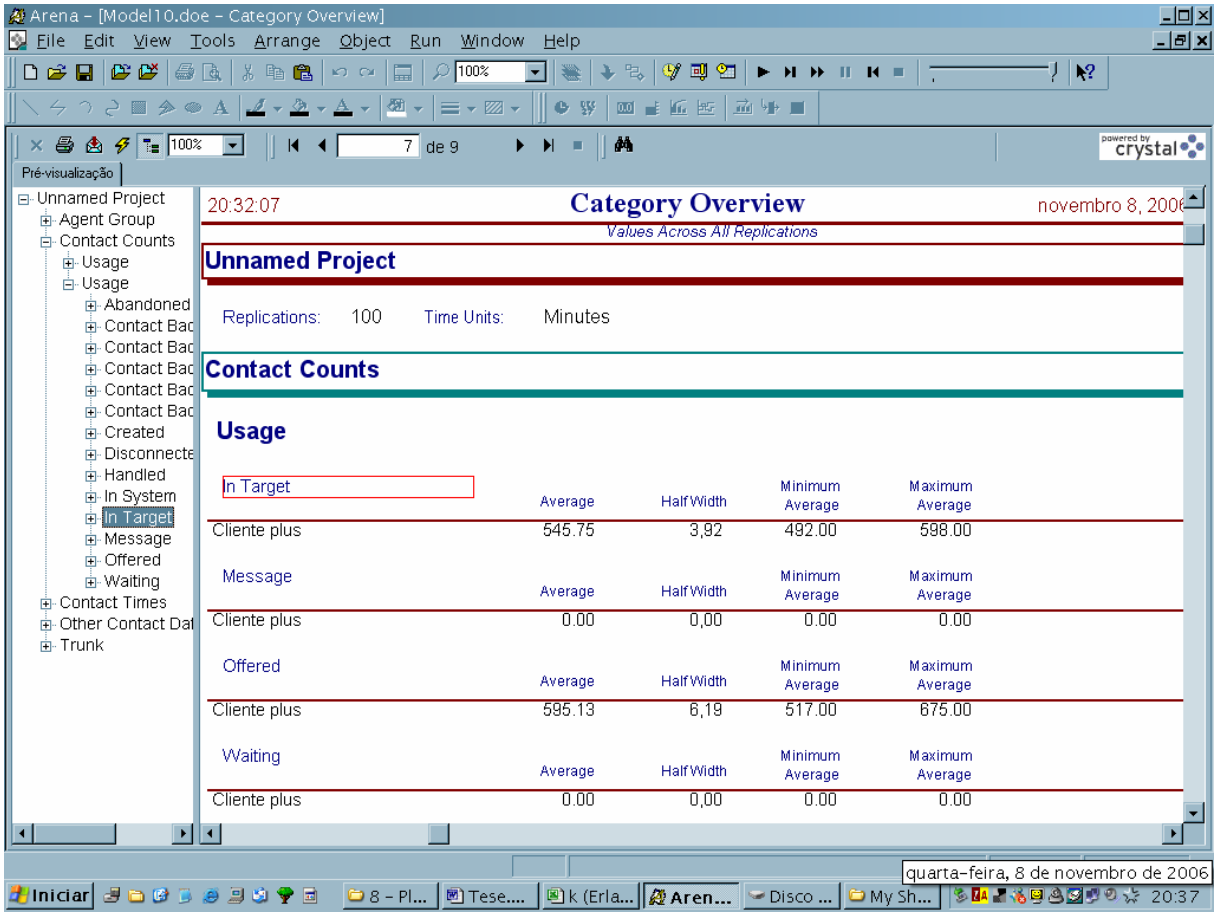
Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

2,8 (0,46%) ligações foram bloqueadas (*blocked*) antes de entrar no sistema, dentre as quais 2,3 tentaram retornar posteriormente (*contact back blocked*), em média; ou seja, houve um retorno de 81,29% das ligações bloqueadas, percentual próximo do valor estabelecido na premissa (85%).

As ligações abandonadas perfizeram 14,03, em média, o que representa 2,35% do total das ligações geradas, cifra muito parecida com a taxa de abandono original (2,44%), o que revela um impacto insignificante da limitação de linhas neste indicador de performance.

Para o cálculo do outro indicador de performance (nível de serviço), é necessário resgatar uma informação a partir da figura 74 a seguir.

Figura 74 – Chamadas completadas e atendidas em até 10 segundos, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

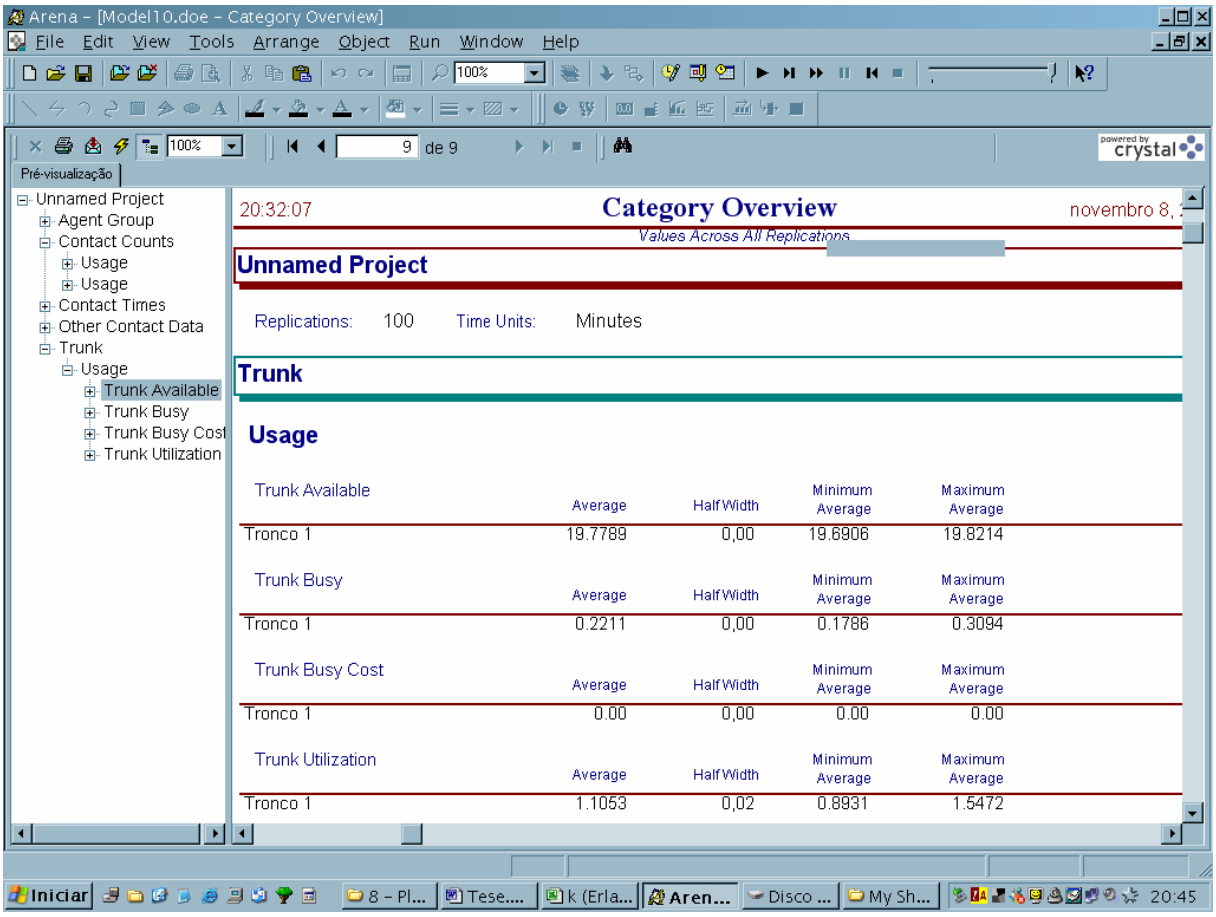
Em média, 546 chamadas foram atendidas em até 10 segundos, ocasionando um nível de serviço de 94,05%. Esse valor – assim como aconteceu com a taxa de abandono – é também próximo dos 93,31% originais.

Fica a impressão de que a limitação na quantidade de linhas não gera impactos significativos nos indicadores de performance. No entanto, um novo número comprometedor da qualidade de serviço (taxa de ligações bloqueadas; no exemplo, 0,46%) apareceu e deve ser levada em conta, ou na forma de um novo medidor de performance ou sendo incorporado (somado) à taxa de abandono (embora o elemento causador da deterioração da qualidade do serviço nesse caso seja diferente do que o que ocorre quando do abandono das chamadas, funcionando como uma espécie de "abandono instantâneo e compulsório").

Na figura 74 anterior, foi possível observar uma grandeza que adquiriu relevância com a restrição na quantidade de linhas: a quantidade de chamadas completadas (ou que conseguiram entrar no sistema). No exemplo anterior, 595 chamadas, em média, atingiram esse *status* (*offered*). A diferença entre esse valor e a quantidade média de chamadas geradas consiste justamente no número médio de ligações que não conseguiram entrar no sistema – as bloqueadas.

A introdução da limitação na quantidade de linhas disponíveis pode criar interesse para o acompanhamento do nível de utilização do tronco de linhas. A figura 75 a seguir apresenta os subsídios para tal.

Figura 75 – Disponibilidade, ocupação e utilização de linhas, das 00:00 às 00:30, agosto de 2006, produto 102 (*plus*), 12 atendentes, cenário com quantidade limitada (20) de linhas disponíveis, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Como pode ser observado, 19,78 das 20 linhas ficaram disponíveis (*trunk available*), em média, ao longo do período, ficando, por conseguinte, apenas 0,22 linhas ocupadas (*trunk busy*) durante o mesmo, em média. O valor baixo (1,11%) da conseqüente taxa de utilização (*trunk utilization)* média revela uma grande ociosidade do tronco, como um todo, fenômeno bastante comum, no entanto, nas centrais de atendimento mais modernas.

Além do impacto da modificação de certos parâmetros do modelo no desempenho do sistema em determinado horário, pode ser interessante – já que uma melhor acurácia no cálculo do nível de serviço foi obtida pela simulação – querer indagar se o mesmo ocorreria em horários com saldos de agentes dimensionados substancialmente diferentes; ou seja, em momentos com grande déficit ou com grande superávit de oferta de capacidade em relação à demanda.

Para o dimensionamento dos atendentes no intervalo entre 05:30 e 06:00 do mesmo mês, foi considerado que chegariam à central 253 chamadas, com um TMA de 31 segundos, conforme pode ser visto na tabela 10 anterior.

De acordo com as fórmulas analíticas implementadas na mesma tabela, seria necessário designar 7 atendentes (já considerando o absenteísmo) para este período, com o intuito de alcançar a meta de nível de serviço para os clientes *plus*.

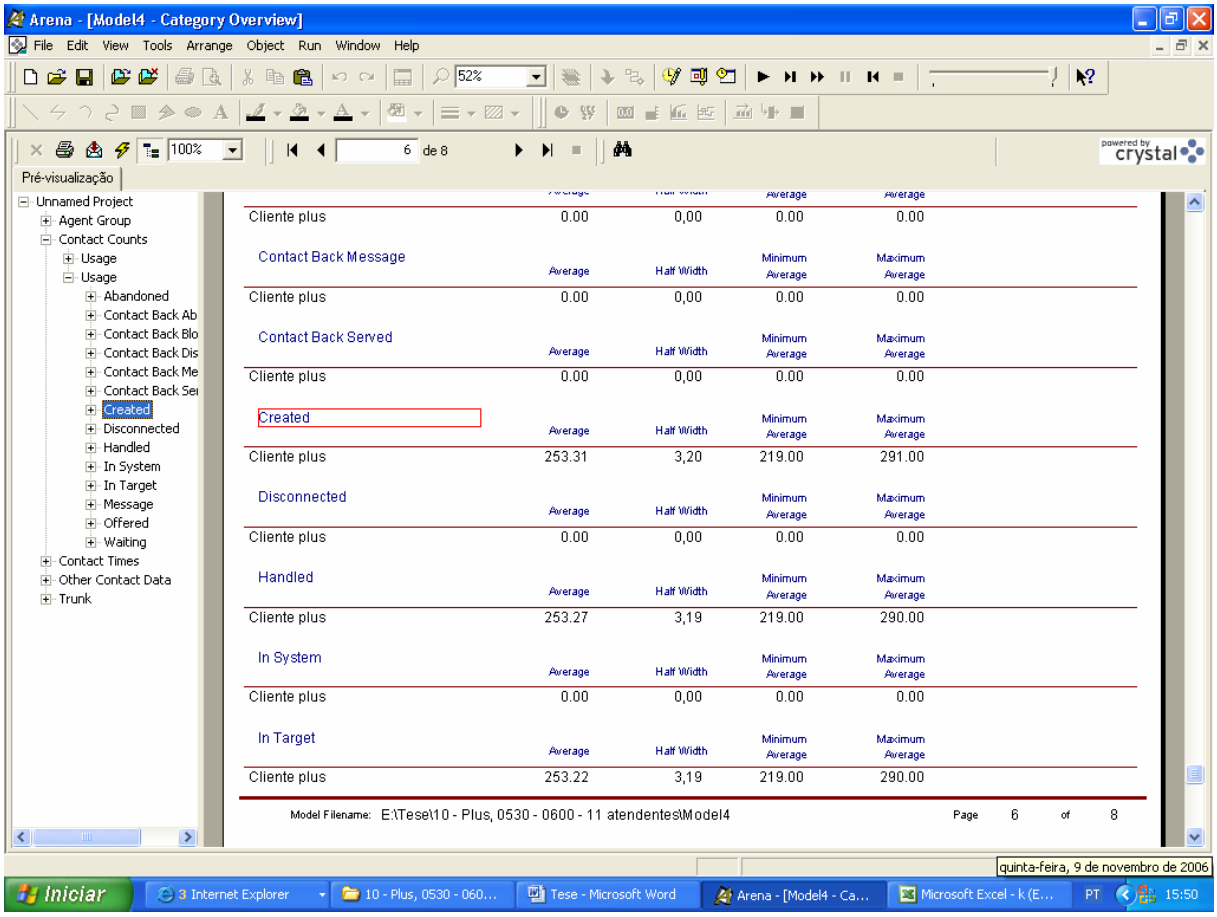
Foram dimensionados 11 agentes (saldo positivo de 4) para este horário e as fórmulas Erlang previram 99,78% de nível de serviço para este horário, já considerando o absenteísmo, que também não influiu neste caso, já que 11 * (1 – 4%) = 11 agentes, aproximadamente.

Com o objetivo de questionar esse nível de serviço estimado, foi construído um modelo com as mesmas premissas de demanda (volume e TMA) e com a mesma capacidade efetiva dimensionada (11 atendentes), desta vez para este segundo horário.

A chegada das ligações e o tempo de atendimento foram modelados de acordo com estas premissas médias e seguindo as distribuições exponencial e Erlang (k = 4), respectivamente; e o abandono foi tratado de forma idêntica à anterior.

A simulação da operação durante este intervalo foi replicada 100 vezes e o *output* pode ser visualizado na figura 76 a seguir.

Figura 76 – Chamadas geradas, atendidas e atendidas em até 10 segundos, das 05:30 às 06:00, agosto de 2006, produto 102 (*plus*), 11 atendentes, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Em média, 253,31 chamadas foram geradas (*created*), 253,27 atendidas (*handled*) e 253,22 em até 10 segundos (*in target*) em cada replicação. Quase não houve abandono, o que não é surpreendente em vista do super-dimensionamento empreendido para este cenário. O nível de serviço resultante foi 99,98%, um valor altíssimo (inclusive quando comparado com a meta de 85%), e bem parecido com os 99,78% previstos pelas fórmulas Erlang para este cenário.

A base de dados da Contax revela que, no dia 15 de agosto de 2006, 11 agentes estiveram atendendo durante o período das 05:30 às 06:00, quando 249 chamadas foram atendidas, com TMA de 31,5 segundos. Estes números estão muito próximos das premissas de demanda usadas nos modelos analítico e de simulação.

Dentro de tal período, foram atendidas 248 chamadas em até 10 segundos (do total de 249), implicando em um nível de serviço real de 99,60%. Desta vez, os valores previstos por ambas as abordagens (analítica e experimental) ficaram muito parecidos com o valor real, até porque não havia muito espaço para grandes variações, visto que este cenário envolve um patamar de nível de serviço já muito próximo de 100%.

A conclusão aponta para a inexistência de ganho de acurácia na previsão do nível de serviço – em cenários com patamares muito altos para essa grandeza – promovido pelo uso da abordagem experimental. Resta saber o comportamento desta acurácia em situações com níveis de serviço muito baixos (déficit de agentes dimensionados em relação à necessidade).

Para o dimensionamento dos operadores (também para agosto de 2006) no período entre 02:00 e 02:30, foi usada a premissa de que chegariam à central 196 chamadas, com um TMA de 28 segundos, como pode ser visto na tabela 10 anterior.

De acordo com a mesma, seria necessário (segundo as fórmulas analíticas) alocar 5 agentes (já considerando o absenteísmo) para este horário, de forma a atingir 85% de nível de serviço.

Foram dimensionados 4 atendentes para este período e as fórmulas analíticas previram 44,05% de nível de serviço para este horário, já levando em conta o absenteísmo, que acabou não influindo neste caso: 4 * (1 – 4%) = 4 atendentes, aproximadamente.

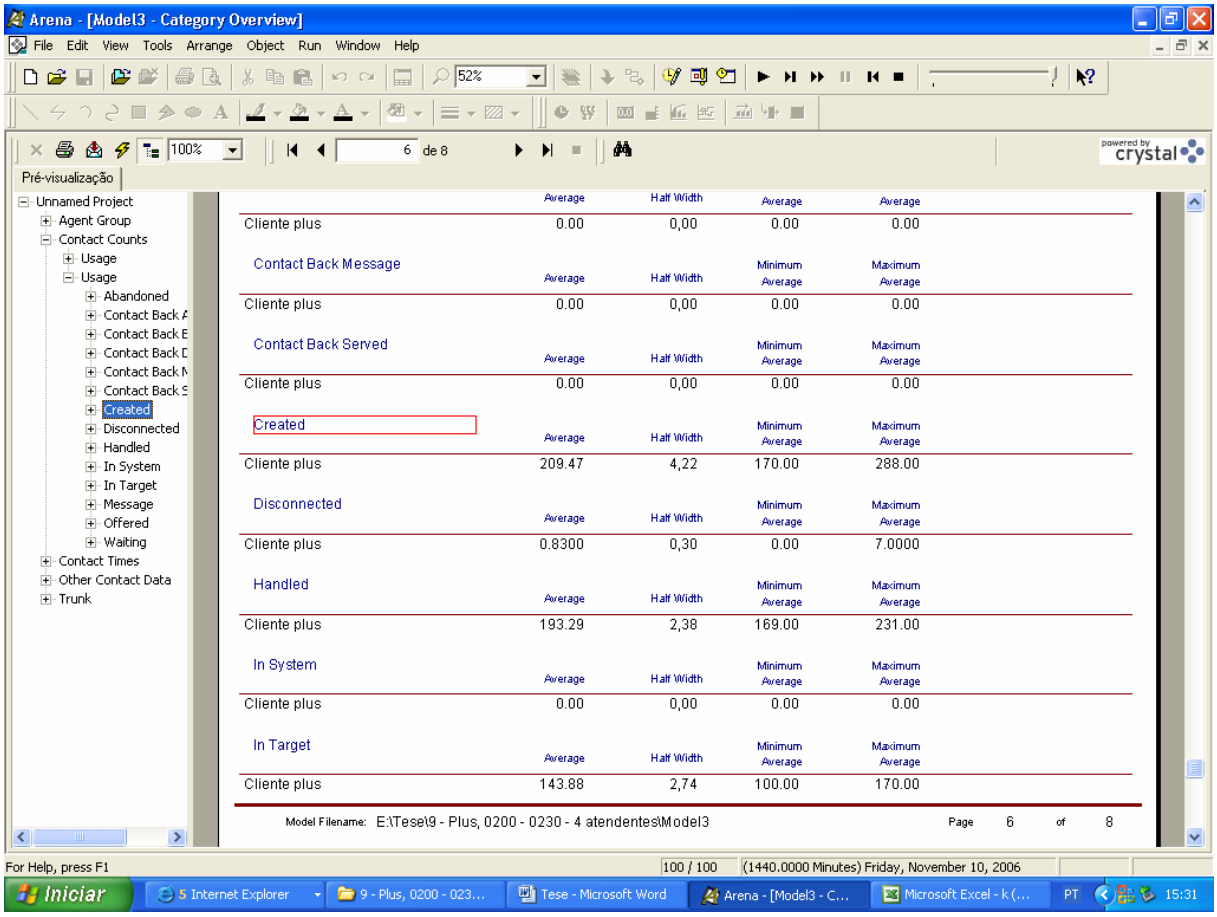
O déficit absoluto de agentes (-1), nesse caso, foi o mesmo do ocorrido no horário entre 00:00 e 00:30; mas, proporcionalmente, ter 1 atendente a menos quando eram necessários 5 é muito mais impactante do que ter 1 a menos quando eram necessários 13. O baixo nível de serviço estimado para este segundo caso (44,05%) ratifica essa afirmação.

Novamente foi construído um modelo, desta vez para simular o comportamento do sistema neste terceiro horário, com as mesmas premissas de demanda (volume e TMA) e com a mesma capacidade efetiva dimensionada (4 agentes).

A chegada das ligações e o tempo de atendimento foram modelados de acordo com estas premissas médias e seguindo as distribuições exponencial e Erlang (k = 4), respectivamente; e o comportamento de abandono modelado da mesma forma que anteriormente.

A simulação destes 30 minutos de operação do *call center* foi replicada 100 vezes e o resultado pode ser observado na figura 77 a seguir.

Figura 77 – Chamadas geradas, atendidas e atendidas em até 10 segundos, das 02:00 às 02:30, agosto de 2006, produto 102 (*plus*), 4 atendentes, 100 replicações



Fonte: Relatório desenvolvido pelo *software* Arena Contact Center

Em média, 209 chamadas foram geradas (*created*) em cada replicação; destas, 193 – em média – foram de fato atendidas (*handled*) pelos operadores em cada replicação, das quais 144 em até 10 segundos (*in target*). O nível de serviço obtido foi 74,44%, um valor mais baixo que a meta de 85%, mas muito acima dos 44,05% previstos pelas fórmulas analíticas que, de fato, parece ter subestimado esta grandeza.

Segundo a base de dados da Contax, no dia 29 de agosto de 2006, 4 operadores atenderam 192 ligações, entre 02:00 e 02:30, com TMA de 27,6 segundos.

Estes números são bem próximos das premissas de demanda utilizadas no dimensionamento através de fórmulas Erlang e neste modelo de simulação. Nesse intervalo, 136 ligações foram atendidas em até 10 segundos (dentre as 192), gerando um nível de serviço real de 70,83%.

Assim como quando aconteceu no cenário original (horário das 00:00 às 00:30), este valor ficou mais próximo (desta vez ainda mais) do nível de serviço calculado pela simulação. Isso tudo revela um ganho ainda maior de acurácia na previsão do nível de serviço, em cenários com baixos patamares para essa grandeza.

O comportamento do ganho de acurácia – empreendido pelo uso da abordagem de simulação – em função do patamar onde se encontra o nível de serviço (que foi investigado ao longo desta seção) pode ser resumido na tabela 41 a seguir.

Tabela 41 – NS real e estimado por fórmulas Erlang e simulação, erros de estimação e ganho de acurácia em diferentes patamares de nível de serviço

Horário	02:00 às 02:30	00:00 às 00:30	05:30 às 06:00
Saldo/déficit de agentes	1 a menos em 5	1 a menos em 13	4 a mais em 7
	-20%	-8%	+57%
Nível de serviço real	70,83%	92,74%	99,60%
Fórmula *Erlang*	44,05%	88,04%	99,78%
Erro	*26,78%*	*4,70%*	*0,18%*
Simulação	74,44%	93,31%	99,98%
Erro	*3,61%*	*0,57%*	*0,38%*
Diferença entre os erros (ganho de acurácia)	**23,17%**	**4,13%**	**-0,20%**

Fonte: Tabela elaborado pelo pesquisador

Este último horário simulado apresenta o nível de serviço em um patamar relativamente baixo, possibilitando uma análise de sensibilidade mais completa do impacto de uma oferta maior de atendentes, conforme mencionado anteriormente nesta seção.

A mesma simulação foi repetida algumas vezes no *software*, sempre para este horário e com os mesmos parâmetros, mas aumentando a quantidade de atendentes. Os resultados mais importantes foram coletados, a partir dos quais os principais indicadores de desempenho (nível de serviço – percentual de ligações atendidas em até 10 segundos – e taxa de abandono) puderam ser obtidos, estando apresentados na tabela 42 a seguir.

Tabela 42 – Nível de serviço e taxa de abandono para diferentes quantidades de atendentes, horário das 02:00 às 02:30, agosto de 2006, produto 102 (*plus*), 100 replicações

Atendentes	Ligações				Nível de serviço	Taxa de abandono
	geradas	atendidas	em até 10 seg	abandonadas		
4	**209**	**193**	**144**	**15,35**	**74,44%**	**7,33%**
5	201	195	180	5,47	92,10%	2,73%
6	195	194	190	0,95	98,17%	0,49%
7	196	195	194	0,62	99,47%	0,32%

Fonte: Tabela elaborada pelo pesquisador a partir de resultados obtidos pelo *software* Arena Contact Center

Como pode ser observado, a alocação de um quinto atendente melhora sensivelmente os indicadores de desempenho, deixando-os com valores bem aceitáveis (principalmente o nível de serviço). A inclusão do sexto agente pode ser considerada interessante, pois encerra uma melhora significativa para a taxa de abandono, mas nem tanto para o nível de serviço, que já era mais do que suficiente. O sétimo atendente parece ser totalmente dispensável, já que as melhorias proporcionadas por ele seriam apenas marginais.

A decisão mais refinada de quantos agentes efetivamente alocar para este horário passaria pela consideração dos potenciais custos envolvidos na inclusão de 1 ou mais atendentes no plano de horários. Dessa forma, a Contax poderia questionar se estaria disposta a gastar mensalmente um determinado montante adicional com mão-de-obra (passando de 4 para 5 agentes) para poder melhorar o nível de serviço e a taxa de abandono no horário em questão de 74% para 92% e de 7,33% para 2,73%, respectivamente.

Em caso afirmativo, o mesmo questionamento poderia ser feito para que pudesse ser verificada a viabilidade da inclusão do sexto e do sétimo agentes, só que cada vez com melhorias menores nos indicadores de desempenho.

Essa análise ratifica a grandeza do impacto da modificação da quantidade de atendentes no desempenho do sistema e a conseqüente importância de um correto dimensionamento da capacidade de atendimento, ambas verificadas anteriormente nesta seção.

7. Conclusões

A indústria de *call centers* tem crescido demasiadamente nas últimas décadas no mundo e de forma bem intensa no Brasil, durante os últimos anos. O mercado doméstico já vem movimentando cifras financeiras bem acentuadas, assim como em outros países. Boa parte dos custos envolvidos na administração das centrais de atendimento é oriunda dos recursos humanos. Por isso, é vital – do ponto de vista da competitividade do negócio – gerenciá-los eficientemente.

Paralelo a esse crescimento, o mercado em questão vem experimentando um aumento muito grande na complexidade das suas operações, dificultando a sua administração através de posturas amadoras e exigindo, portanto, tomadores de decisão e ferramentas cada vez mais profissionais.

Mas, principalmente no Brasil, muito pouco material acadêmico foi produzido a respeito de técnicas e ferramentas capazes de apoiar a tomada de decisão gerencial, no tocante à administração dos recursos, no contexto dos *call centers*. Torna-se clara, então, a necessidade de descobrir – entre outras inúmeras facetas deste problema, a serem atacadas por outros pesquisadores – quais são as melhores metodologias para abordar e resolver alguns dos principais desafios e problemas quantitativos relacionados à administração de *call centers*.

Antes, naturalmente, foi preciso levantar quais são estes desafios, quando as questões de previsão de demanda e de dimensionamento da capacidade de atendimento apareceram como as principais respostas.

Em relação ao problema de previsão da demanda de ligações, foi necessário pesquisar algumas metodologias que costumam ser utilizadas, assim como investigar a sua adequação às características da indústria e propor soluções melhores para lidar com este desafio.

Já no que diz respeito ao problema de dimensionamento da capacidade de atendimento, formou-se uma dicotomia, opondo métodos analíticos (como a teoria de filas) a métodos experimentais (como a simulação).

E já que a simulação, geralmente, é uma ferramenta mais cara e de implementação mais difícil e custosa, foi preciso investigar – no cenário de *call centers* – quais são os *trade-offs*

envolvidos na decisão de qual metodologia usar e como tal abordagem pode ajudar os gerentes das centrais no tratamento deste desafio operacional.

Desta forma, este trabalho procurou, inicialmente, introduzir o tema, contextualizar o problema, apresentar as principais questões envolvidas e as características do estudo. Em seguida, o setor e a empresa estudados foram caracterizados e foi feita uma revisão de literatura acerca dos principais temas abordados. O trabalho prosseguiu apresentando seus aspectos metodológicos, incluindo a descrição de como os dados foram coletados e tratados, de forma a atingir os objetivos propostos.

Na seqüência, esta pesquisa objetivou descrever o caso desenvolvido para dar resposta às questões propostas, apresentando o setor da empresa envolvido, além de seus principais desafios operacionais, que foram explorados na última parte deste texto, quando foram apresentadas soluções concretas para os mesmos através de "novas" (do ponto de vista da empresa) metodologias, que foram desenvolvidas e ajustadas para atacar os problemas em questão.

A seguir, as respostas às questões propostas são apresentadas, também separadas em grupos, conforme foi feito na seção 1.1.2.

7.1. Respostas às questões propostas

7.1.1. Previsão

a) Que metodologia(s) é(são) aplicada(s)?

Foi possível constatar, no decorrer do trabalho, que o problema de previsão de demanda costuma ser tratado pela Contax, de uma forma geral, através da ferramenta de previsão que o pacote computacional utilizado na empresa (TotalView) fornece – uma espécie de amortecimento exponencial com a inclusão da sazonalidade intra-semanal – além de alguns ajustes subjetivos promovidos pela equipe de planejamento.

b) Essa(s) é(são) a melhor forma de prever tais grandezas?

O problema é que esta ferramenta de previsão não leva em conta algumas informações facilmente acessíveis e que poderiam subsidiar melhor as predições. Assim, estas informações

não são incluídas de forma sistemática no processo de previsão, ficando essa inclusão um pouco por conta da subjetividade (carregada de uma eventual arbitrariedade) dos analistas.

Definitivamente, ao menos no caso do produto 103 (serviços de telefonia fixa e problemas com a conta telefônica), não incluir de forma sistemática informações referentes à ocorrência de feriados e às datas de vencimento e chegada das contas na residência dos clientes consiste em desperdício de informação muito útil e de fácil acesso.

c) Se não for(em), qual é?

Foi desenvolvido um modelo de regressão múltipla (com variáveis *dummy*) para prever o volume diário de ligações, assim como o seu tempo médio de atendimento (TMA), levando em consideração – de forma sistemática – as informações acima citadas, além da sazonalidade intra-semanal (já levada em conta pelo TotalView).

As previsões elaboradas pelo modelo de regressão foram comparadas aos dados reais de demanda e revelaram uma acurácia um tanto superior à obtida pela ferramenta de previsão do TotalView.

7.1.2. Dimensionamento

d) Qual metodologia é aplicada para a resolução desse problema?

Em relação a este problema, a Contax utiliza uma metodologia analítica (fórmulas Erlang) para calcular o nível de serviço e monitorar indiretamente o abandono de ligações (principais indicadores de performance da operação) em cada cenário de dimensionamento de efetivo proposto pela equipe de planejamento.

e) A simulação é utilizada?

A simulação é praticamente desconhecida – enquanto ferramenta de apoio ao dimensionamento – para os analistas e, até por isso, não é utilizada para este fim.

f) Se não for, por que não é?

Além do pequeno conhecimento por parte dos analistas de dimensionamento acerca da abordagem, as aparentes complexidade e grande exigência de *inputs* da metodologia também constituem obstáculos para a sua utilização, na opinião da equipe de planejamento.

g) Quais são os *trade-offs* envolvidos na decisão de qual metodologia usar (simulação *versus* outras)?

Foi possível constatar, ao longo dessa pesquisa e através do uso de modelos de simulação para lidar com o problema de dimensionamento da capacidade de atendimento do *call center*, diversas vantagens da abordagem experimental em comparação às analíticas, especialmente em operações mais complexas:

- é possível incluir mais detalhes da operação (conforme será especificado na resposta à questão "i"), utilizar distribuições estatísticas mais condizentes com os dados de entrada e deixar o modelo mais próximo da realidade, garantindo a obtenção de resultados mais acurados;

- o nível de serviço calculado pelas fórmulas Erlang normalmente é subestimado, principalmente pelo fato destas ignorarem o abandono de chamadas e utilizarem o formato exponencial para a distribuição do tempo de atendimento;

- outros indicadores de desempenho (não disponíveis nas abordagens analíticas, como a taxa de abandono e o percentual de clientes bloqueados) podem ser avaliados, apresentados e estudados;

- valores mínimo e máximo de cada indicador importante podem ser obtidos, não ficando o analista restrito à media dos valores, como no caso da teoria de filas;

- um entendimento melhor da operação é atingido com a adoção da abordagem experimental, onde existe a possibilidade de acompanhar dinamicamente o comportamento do sistema e dos seus indicadores de performance e, dessa forma, entender por que as filas estão se formando e o motivo dos tempos de espera estarem altos, por exemplo, ao passo que, através da metodologia Erlang, é possível enxergar apenas os *outputs* gerados (indicadores numéricos) em função dos *inputs* fornecidos, dificultando o entendimento completo da operação;

- a comunicação pode tornar-se mais fácil, através de animações gráficas.

No entanto, uma eventual implementação desta metodologia experimental demanda, sem dúvida, um intenso esforço, alem de acarretar custos elevados (de aquisição do *software* e de treinamento) e uma necessidade de conhecimento técnico.

h) Como a simulação pode dar apoio às decisões referentes ao processo de dimensionamento do *call center*?

Ao longo do estudo de caso realizado a respeito da operação de atendimento dos clientes do produto 102 (auxílio à lista), diversos modelos de simulação foram construídos – contemplando diferentes características reais dos *call centers* e para diferentes possíveis cenários alternativos – para calcular indicadores de performance e sugerir soluções para o dimensionamento da operação.

O cálculo de um indicador em especial – a taxa de abandono – pode ser extremamente útil, no caso da Contax. Isso ocorre porque, conforme descrito na seção 5.2.2, os analistas de dimensionamento baseiam muito suas decisões no déficit resultante de atendentes em cada intervalo de meia-hora. Eles procuram manter essa grandeza em patamares baixos para evitar que haja uma alta incidência de abandono.

No entanto, o ideal seria que eles pudessem avaliar – diretamente – a taxa de abandono ocasionada por cada alternativa de escalonamento; mas como a abordagem analítica por eles utilizada desconsidera o abandono, não permitindo a quantificação da sua ocorrência, só lhes resta a sua avaliação indireta através mesmo do déficit de agentes, estando essa sujeita, obviamente, a altas doses de subjetividade e intuição. A simulação acaba com esse problema, já que a sua utilização permite a estimação direta e acurada da taxa de abandono para cada sugestão de escalonamento, balizando as conclusões e tornando-as livres da dependência exclusiva da intuição e da subjetividade.

De uma forma geral, ficou claro que a ferramenta em questão permite avaliar, com facilidade, o impacto – nos indicadores de performance – de modificações nas características originais da operação (como, por exemplo, uma redução ou um aumento de oferta de capacidade de atendimento ou uma redefinição do conceito interno de nível de serviço) e realizar análises de sensibilidade completas acerca de alguns parâmetros operacionais.

Especificamente, as análises de cenários e de sensibilidade desenvolvidas durante o estudo de caso permitiram visualizar como a simulação pode dar apoio às decisões referentes ao

processo de dimensionamento da central, uma vez que os resultados revelaram que, entre outras:

- pode ser possível diminuir o efetivo de atendentes em alguns horários do dia, sem comprometer a meta de nível de serviço, já que o cálculo desta grandeza empreendido pelas fórmulas Erlang normalmente a subestima;

- mas reduções deste efetivo para valores abaixo do mínimo indispensável (evidenciado para um horário específico na análise de sensibilidade conduzida) podem trazer conseqüências catastróficas para a operação do sistema, especialmente em termos da taxa de abandono;

- por outro lado, a utilização de um efetivo de atendimento maior pode melhorar sensivelmente a performance do sistema (principalmente em termos de nível de serviço), mas apenas até um certo limite (também evidenciado para um horário específico na análise de sensibilidade conduzida), acima do qual alocações adicionais de operadores seriam capaz de proporcionar apenas melhorias marginais na qualidade do atendimento;

- variações não muito grandes no volume de ligações recebidas podem impactar bastante os indicadores de desempenho, especialmente a taxa de abandono, apontando para a necessidade de se prever tal grandeza com bastante acurácia;

- esses impactos são ainda maiores no caso de variações no tempo médio de atendimento, tornando a previsão desta variável ainda mais crítica e – já que o comportamento desta grandeza não pode ser considerado totalmente alheio à vontade da gerência – indicando que pode ser interessante para a empresa investir – até um certo limite – em treinamento e conscientização para a redução desse tempo;

- a variabilidade do tempo de atendimento também é capaz de impactar de forma significativa os indicadores de desempenho, especialmente o nível de serviço, mostrando que é importante tentar mensurar também o desvio-padrão daquela grandeza (e não apenas a sua média) e mantê-la homogênea (estável) através de treinamento e conscientização;

- é possível melhorar consideravelmente a performance de atendimento para os clientes preferenciais, sem comprometer demasiadamente a qualidade do serviço para os clientes comuns, se for adotado um formato agregado – com prioridades – de atendimento;

- na hipótese dos clientes se tornarem mais impacientes e abandonarem as ligações mais rapidamente, o impacto no nível de serviço seria pequeno, mas bem significativo no caso da taxa de abandono – revelando que a importância de determinadas mudanças de cenário pode depender dos indicadores de desempenho utilizados e tidos como importantes pela empresa;

- na hipótese da central dispor de uma quantidade limitada (e pequena) de linhas disponíveis no tronco, o impacto nos indicadores de performance tradicionais (nível de serviço e taxa de abandono) seriam pequenos, mas seria necessária a criação de um outro indicador (percentual de clientes bloqueados) capaz de capturar essa deterioração no nível de atendimento da demanda.

i) Que aspectos operacionais do *call center* não podem ser bem tratados por outras metodologias, sendo determinantes para justificar o uso da simulação?

O *call center* revelou-se uma operação bastante complexa, ficando claro que alguns de seus aspectos e características não podem ser adequadamente explorados por abordagens analíticas (justificando o uso de uma ferramenta experimental elaborada – como a simulação – para lidar com o problema de dimensionamento das centrais):

- comportamento de abandono por parte dos clientes;

- disponibilidade limitada de linhas telefônicas no tronco;

- variedade de clientes, com diferentes comportamentos e prioridades de atendimento;

- agentes com diferentes níveis de habilidade e preferências nas atividades de atendimento;

- diferentes disciplinas de fila;

- desvios e transferência de chamadas;

- conferências envolvendo o cliente e múltiplos atendentes.

j) A simulação fornece resultados mais acurados em algumas situações do cotidiano operacional do *call center*? Em quais situações?

A maior acurácia dos resultados é obtida nas centrais mais complexas operacionalmente, especialmente em situações envolvendo grandes incertezas quanto à demanda e aos tempos de atendimento.

O estudo em questão permitiu concluir que o ganho de acurácia para o nível de serviço promovido pela simulação é maior nos horários onde essa grandeza se encontra em patamares mais baixos, conforme a tabela 41 anterior.

k) Quais *softwares* estão disponíveis para o uso da simulação aplicada a *call centers*?

A pesquisa revelou uma série de *softwares* disponíveis para a aplicação da simulação em *call centers*: Arena Contact Center; callLAB; SimACD; Service Model; Simul8; ccProphet; Contact Centers; Excel Plataform for Simulating Call Centers; IVR Simulator Model.

l) Quais são as suas características?

Os quadros comparativos apresentados nas tabelas 2 e 3 anteriores procuraram sumarizar e comparar as características gerais de cada *software* e os atributos das operações de *call centers* que cada um é capaz de contemplar.

O primeiro quadro revelou, entre outras coisas, que a maioria das opções empreende a terminologia própria de *call centers* e permite a observação dinâmica dos indicadores de performance, mas quase nenhuma possui alguma ferramenta para fazer um tratamento estatístico preliminar dos dados de entrada.

O segundo quadro mostrou que quase todos os *softwares* contemplam o comportamento de abandono por parte do cliente. A maioria também permite a escolha da distribuição dos tempos de atendimento e entre chegadas. Poucos *softwares*, no entanto, são capazes de considerar algumas características das centrais, como: a "re-tentativa" de ligação por parte do cliente, tele-conferência e transferência de chamadas.

Entre as opções, o Arena Contact Center revelou-se como o mais completo dos *softwares*, não apresentando, no entanto, algumas características gerais importantes, como simplicidade e flexibilidade.

m) Quais deles são utilizados no *call center* estudado?

Já que, na Contax, a simulação não é utilizada como metodologia para lidar com o problema de dimensionamento, nenhum destes *softwares* é utilizado na empresa.

7.2. Sumário das contribuições da pesquisa

A partir das respostas às questões propostas, é possível notar que as principais contribuições desta pesquisa originaram-se do desenvolvimento e da aplicação de "novas" metodologias aos problemas da empresa estudada, podendo ser destacadas, de forma bem sucinta, como:

- a consideração – de forma sistemática – das informações referentes às datas de vencimento e chegada das contas na residência dos clientes no processo de previsão de demanda de ligações para o produto 103;

- a revelação de que resultados a simulação é capaz de mostrar e de que respostas ela pode fornecer para auxiliar no processo de dimensionamento da capacidade de atendimento de um *call center*, destacando as suas vantagens em relação às abordagens analíticas e enfatizando quais peculiaridades das operações da indústria em questão são responsáveis por corroborar a sua adequação.

7.3. Sugestões e recomendações

Infelizmente, foi inviável – do ponto de vista operacional – explorar todos os assuntos pertinentes. Esta pesquisa não teve a pretensão, então, de atacar mais do que apenas uma amostra das questões que poderiam ter sido trabalhadas dentro do tema principal. A seguir, estão relacionados alguns tópicos correlatos que ficaram de fora dessa amostra, consistindo em potenciais objetos de estudo para futuras pesquisas.

7.3.1. Ferramentas de previsão de demanda

A consideração de regionalidades por parte dos clientes seria um aspecto interessante a ser tratado em pesquisas futuras sobre previsão de demanda em *call centers*. Por exemplo, o impacto na demanda causado pelo fato de uma data específica consistir do dia seguinte à chegada da conta telefônica para 200.000 clientes do Rio de Janeiro pode ser diferente do impacto proporcionado pelo fato da data consistir do dia seguinte à chegada da conta para 200.000 clientes da Bahia. Seria interessante a elaboração de um modelo capaz de levar em conta uma eventual diferença entre o comportamento cultural dos cariocas e baianos em relação à chegada da conta; e que não tratasse a demanda de forma consolidada nacionalmente e nem considerasse os impactos de forma homogênea.

Outra sugestão acerca de uma ferramenta de previsão de demanda consiste na elaboração de um modelo capaz de capturar, entender e prever os efeitos de uma possível sazonalidade ao longo do ano.

Ainda a respeito de tal ferramenta, pesquisadores interessados poderiam tentar desenvolver um modelo para fazer a previsão intradiária da demanda de ligações, já que a maioria dos *call centers* utiliza os blocos de 30 minutos como unidade básica de análise (para previsão e dimensionamento), tornando a predição em curto prazo da quantidade de chamadas muito importante.

A atual pesquisa revelou um comportamento inesperado do tempo médio de atendimento (TMA) em função da chegada das contas telefônicas à residência dos clientes: na verdade, uma ausência de dependência, o que surpreende pelo fato da quantidade de ligações ter sido identificada como dependente de tal fator. Como esta grandeza (o TMA) é fundamental para o correto dimensionamento das equipes de atendentes nas centrais, faz-se necessária uma pesquisa quantitativa exploratória para tentar explicar esse comportamento inusitado. O entendimento de tais motivos poderiam acarretar uma melhoria na qualidade de futuras ferramentas de previsão.

Um assunto semelhante pode, da mesma forma, instigar pesquisadores futuros: os motivos que reduzem o tempo médio de atendimento nos feriados e finais de semana (em até 30 segundos, quando comparados aos dias úteis) não foram descobertos e merecem ser investigados por meio de uma pesquisa mais profunda e qualitativa.

Uma última sugestão a respeito de ferramentas de previsão de demanda aponta para a necessidade de extrapolar o uso dos modelos de regressão – como o aqui apresentado – para prever a quantidade de chamadas e o tempo médio de atendimento de outros produtos da própria Contax, como Velox, OI, 102 ou suporte técnico. Em relação a este último, por exemplo, o modelo de regressão poderia usar como variáveis explicativas – assim como o aqui apresentado utilizou o dia da semana, a ocorrência de feriados e a proximidade com dias críticos em relação às contas telefônicas – fatores razoavelmente previsíveis e impactantes na demanda de ligações acessando o serviço, como a ocorrência prevista de chuvas, entre outros.

7.3.2. Dimensionamento da capacidade de atendimento

Muito tem sido feito desde que se percebeu a complexidade inerente ao processo de gerenciamento de *call centers* modernos. Naturalmente, há ainda muito espaço para serem respondidas – mais e de forma mais acurada – questões relativas à otimização de centrais de atendimento, através de mais discussões e novos estudos, e da melhoria em um aspecto fundamental nos sistemas de modelagem: a proximidade com a realidade.

Uma sugestão para pesquisadores que queiram realizar trabalhos na área de simulação de *call centers* consiste na simulação contínua da operação da central ao longo de, por exemplo, um dia inteiro (mesmo que sejam necessárias diferentes quantidades de agentes em cada horário), com vistas ao completo dimensionamento do efetivo de atendimento em tal horizonte de tempo. A continuidade pode trazer mais acurácia e outros benefícios para os resultados, além de facilitar o processo, já que não seria necessário simular a operação em diversos intervalos pequenos de tempo.

A maioria dos *softwares* de simulação de *call centers* está capacitada a reproduzir a operação por longos períodos, continuamente, bastando que seja utilizada a sua versão completa, que permite o uso de uma quantidade grande de PAs.

Para modelar de forma mais acurada as situações a serem simuladas, seria interessante que pesquisas futuras empreendessem um esforço para descobrir (baseadas no histórico de ligações) a correta distribuição estatística e variabilidade do tempo entre chegadas de ligações e do tempo de atendimento. Parte dos estudos de simulação assume a premissa básica que estes tempos seguem uma distribuição exponencial e realizam pesquisas apenas a respeito da média que estas variáveis apresentam em cada bloco unitário de tempo. Mas os resultados

obtidos podem ser sensíveis ao formato da distribuição e à variabilidade destes tempos, de acordo com o que foi pesquisado durante a revisão bibliográfica.

Um outro objeto de pesquisa interessante seria o mapeamento – e a inclusão no modelo – do comportamento pós-abandono dos clientes: a parcela percentual que tenta ligar novamente e o formato da distribuição do tempo transcorrido até a próxima tentativa.

Esforço semelhante também pode ser válido no sentido de tentar entender o comportamento de retorno da ligação após a chamada ser bloqueada (antes de entrar no sistema) por falta de linhas disponíveis no tronco. A correta consideração desses comportamentos de retorno pode gerar impactos nos resultados obtidos pelos modelos.

Adicionalmente, poderia também ser levado em consideração o fato de que, após abandonar a ligação inicialmente, o cliente que volta a tentar contato pode estar mais impaciente e acabar aguardando menos tempo desta vez, antes de abandonar novamente.

Ainda dentro dessa linha, uma pesquisa empírica poderia coletar informações a respeito do impacto no tempo de atendimento ocorrido quando as chamadas não são atendidas pelo tipo de operador habituado a atendê-las, em um sistema de atendimento consolidado para chamadas de diferentes tipos. A consideração correta deste impacto tende a gerar resultados mais acurados para os indicadores de performance em centrais com atendimento agregado de ligações.

Estudos futuros de simulação em *call centers* poderiam considerar o processo de transferência da chamada durante o atendimento a um cliente, até que este seja servido pelo funcionário correto. Normalmente, este processo piora a satisfação do cliente, tornando ideal a segmentação *a priori* na URA.

Não costuma ser interessante deixar o atendente ou mesmo o cliente tomar a decisão de para que local a chamada deve ser direcionada; o direcionamento automático costuma ser a melhor solução. Como boa parte dos *softwares* de simulação contempla a possibilidade de transferência nos seus roteiros, este processo poderia ser explorado em modelos computacionais.

Da mesma forma, trabalhos futuros de simulação na indústria de *call centers* poderiam explorar outras peculiaridades deste tipo de operação, que não costumam ser bem abordadas através das metodologias analíticas, como, entre outras:

- conferências entre o cliente e mais de um atendente ao mesmo tempo;
- desvios condicionais de chamadas para atendimentos especializados;
- outras disciplinas de fila, diferentes da tradicional FIFO.

Já que a maioria dos *softwares* de simulação contempla tais atributos operacionais, não seria complicado – para um pesquisador interessado nestas sugestões e disposto a se aprofundar nos detalhes funcionais da ferramenta por ele utilizada – modelar estas peculiaridades e alcançar conclusões interessantes para a indústria em questão.

Uma pesquisa futura poderia tentar descobrir qual é – para fins de dimensionamento – o ponto de equilíbrio em termos de complexidade em operações de *call center* – ou, em outras palavras, quão complexa uma operação tem que ser para passar a justificar o uso de uma abordagem relativamente cara e com grande exigência de conhecimento técnico. Isto porque talvez nem todos os tipos de operação em *call centers* sejam suficientemente complexos para justificar o uso da simulação no processo de dimensionamento da capacidade de atendimento.

Nesse caso, seria interessante tentar identificar que características operacionais são determinantes para tornar a modelagem suficientemente complexa a ponto de justificar o uso da simulação. E também verificar a adequabilidade de diferentes metodologias (analíticas *versus* experimentais) a operações com diferentes níveis de complexidade.

Para buscar essas respostas, algumas operações de *call centers* teriam que ser divididas em diferentes níveis de complexidade (definidos junto com os responsáveis operacionais) e modeladas em um ambiente de simulação. Soluções ótimas para o dimensionamento seriam sugeridas pela simulação e também por metodologias alternativas (analíticas).

As soluções sugeridas pelas diferentes metodologias seriam efetivamente implementadas no sistema real por um curto período de tempo e os resultados em termos de desempenho do sistema seriam usados como base de comparação. Estes resultados consistiriam na base para a determinação dos benefícios de cada metodologia.

Os custos de cada alternativa seriam compostos não apenas pelos custos financeiros, mas também por outros custos de implementação (tempo, mudança, treinamento etc). Dessa forma, poderia ser montado um quadro comparativo de custo-benefício entre as ferramentas, revelando a sua adequabilidade aos diferentes níveis de complexidade da operação.

Com a adequabilidade avaliada para cada nível de complexidade, seria possível verificar a partir de qual nível de complexidade (ponto de equilíbrio) o uso da abordagem experimental já seria justificável. Finalmente, seriam procuradas – entre as operações avaliadas como adequadas ao uso da simulação – características comuns para que pudesse ser feito um mapeamento a respeito de quais características das operações de *call centers* são responsáveis pela complexidade que justifica o uso da abordagem.

7.3.3. Operador multi-produto

Um outro ponto que revela um grande potencial de melhoria operacional a ser estudado em pesquisas futuras diz respeito ao operador multi-produto. O gerente de Planejamento de Tráfego da Contax tem a impressão que é economicamente vantajoso utilizar um mesmo operador para atender duas ou mais operações diferentes ao mesmo tempo, de modo a tentar reduzir a sua ociosidade. Isso poderia acontecer no caso de operações relativamente similares em que o mesmo operador pudesse trabalhar e que apresentassem comportamentos complementares de demanda ao longo do dia (ou da semana ou do mês).

O trabalho de Cauduro *et al* (2002) apresenta o processo de aprendizagem organizacional no setor de *call center* de uma empresa multinacional de telecomunicações durante a mudança de cultura organizacional ocasionada por um processo de fusão e aquisição. Para acompanhar a dinâmica e as peculiaridades específicas do setor, a empresa adotou em seu *call center* o modelo de atendente universal, no qual os funcionários estão habilitados a atender todos os clientes independentemente dos tipos de serviços (pré-pago, pós-pago, cobrança, habilitação). Essa mudança foi desencadeada pela necessidade de melhorar a competitividade da empresa diante do mercado, visto que seus índices em volume de atendimento apresentavam-se aquém dos padrões mundiais, gerando um serviço insatisfatório para o cliente no que tange a tempo de espera e custos para a empresa.

O trabalho dos autores fala justamente de uma espécie de operador multi-produto (batizado por eles de "atendente universal") e inclui as motivações para a sua implantação. Mas a sua

análise aconteceu sob o ponto de vista de comportamento organizacional e de recursos humanos.

Para abordar essa questão do ponto de vista operacional – se o uso de um operador para diferentes operações é vantajoso em termos de custo-benefício, em vista da redução da sua ociosidade – é necessário o uso da simulação, pois a situação é complexa demais para ser resolvida através de algum método analítico. Inicialmente, é recomendável identificar – dentro de uma empresa fornecedora de serviços de *call center* terceirizados – operações relativamente similares em que o mesmo operador pudesse trabalhar, para então modelá-las e simulá-las.

Durante a modelagem, é necessário levar em conta uma eventual redução na eficiência do operador oriunda da "desespecialização". Os impactos esperados no desempenho do sistema causados pela redução da ociosidade do operador precisam ser coletados. Conseqüências negativas intangíveis desse processo (insatisfação do cliente final por causa de um operador menos eficiente, insatisfação da empresa-cliente por não ter mais um operador exclusivo, dificuldades no treinamento de um mesmo operador em duas ou mais operações tão distintas em termos de complexidade que podem necessitar de treinamento por períodos que variam de 1 semana a 1 mês, entre outras) também têm que ser incluídas – juntamente com o impacto no desempenho do sistema – em um quadro de vantagens e desvantagens da situação testada, para ajudar na resposta acerca da viabilidade do formato.

O caráter claramente intangível e de difícil mensuração destas conseqüências negativas, no entanto, pode conferir um certo grau de arbitrariedade à análise custo-benefício e limitar – portanto – um pouco a qualidade dos resultados a serem obtidos.

8. Referências bibliográficas

ALAM, M. Using Call Centers to Deliver Public Services. ***House of Commons Paper.*** London: The Stationery Office Books, 2002

ALEXANDER COMMUNICATIONS GROUP INC. Call Center Scheduling Is Critical for Your Success: Here's What You Need to Know to Make It Easier. ***Customer Service Newsletter***, Dec 21, 2005

ALVES, M. ***Como Escrever Teses e Monografias.*** Rio de Janeiro: Campus, 2003

ANDREWS, B., CUNNINGHAM, S. LL Bean improves call-center forecasting***. Interfaces*** 25, p. 1-13, 1995

ARAÚJO, M., ARAÚJO, F., ADISSI, P. Modelo para segmentação da demanda de um call center em múltiplas prioridades: estudo da implantação em um call center de telecomunicações***. Revista Produção On Line***, volume 4(3), Agosto 2004

ATLASON, J., EPELMAN, M., HENDERSON, S. Call Center Staffing with Simulation and Cutting Plane Methods. ***Annals of Operations Research***, vol. 127(1-4), p. 333-358, 2002

AVRAMIDIS, A., L'ECUYER, P. Modeling and Simulation of Call Centers. ***Winter Simulation Conference***, 2005

AZEVEDO, M., CALDAS, M. Seriam os call centers os sweatshops do século XXI? ***Encontro de Estudos Organizacionais***, Recife, 2002

BANKS, J., CARSON, J., NELSON, B. ***Discrete-event System Simulation.*** New Jersey: Prentice Hall, 1999

BAPAT, V., PRUITTE Jr, E. Using simulation in call centers. ***Winter Simulation Conference***, p. 1395-1399, 1998

BARD TECHNOLOGIES INC. ***callLAB***. Disponível em: www.bardtech.com. Acesso em: 20 set. 2005

BARBOZA, A., CARNIERI, C., STEINER, M., SIQUEIRA, P. A Pesquisa Operacional na Otimização de Atendentes em Centrais Telefônicas. ***Gestão & Produção***, vol. 10(1), abril 2003

BATT, R., DOELLGAST, V., KWON, H. U. S. Call Center Industry Report 2004 – National Benchmarking Report – Strategy, HR Practices & Performance. ***Center for Advanced Human Resource Studies (Cornell University)***, Working Paper 05 – 06, 2004

BELGE SIMULAÇÃO. ***ServiceModel***. Disponível em: www.belge.com.br/produtos/servicemodel. Acesso em: 20 set. 2005

BETTS, A., MEADOWS, M., WALLEY, P. Call centre capacity management***. International Journal of Service Industry Management,*** vol. 11(2), p. 185-196, 2000

BIANCHI, L., JARRET, J., HANUMARA, R. Forecasting incoming calls to telemarketing centers***. Journal of Business Forecasting Methods and Systems,*** 12, p. 3-12, 1993

BIANCHI, L., JARRET, J., HANUMARA, R. Improving forecasting for telemarketing centers by ARIMA modelling with intervention***. International Journal of Forecasting,*** 14, p. 497-504, 1998

BORST, S., MANDELBAUM, A., REIMAN, M. Dimensioning Large Call Centers***. Operations Research,*** vol. 52(1), p. 17-34, Jan-Feb 2004

BUIST, E., L'ECUYER, P. A Java Library for Simulating Contact Centers. ***Winter Simulation Conference***, 2005

BRASIL está entre os 10 países mais atraentes para call centers. ***Estatísticas***. Disponível em: www.callcenter.inf.br. Acesso em: 20 set. 2005

BRASIL fecha 2003 com mais de 68 mil PAs. ***Estatísticas***. Disponível em: www.callcenter.inf.br. Acesso em: 20 set. 2005

BROWN, L., GANS, N., MANDELBAUM, A., SAKOV, A., SHEN, H., ZELTYIN, S., ZHAO, L. Stastical analysis of a telephone call center: a queueing-science perspective. ***Wharton Financial Institutions Center***, Working Paper, 03 – 12, Nov 9, 2002

CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Serviços. ***Banco de Teses***. Disponível em: www.capes.gov.br. Acesso em: 29 jan. 2006

CAUDURO, F., GRAMKOW, F., CARVALHO, M., RUAS, R. O Processo de Mudança e Aprendizagem no Call Center de uma Empresa de Telecomunicações. ***EnANPAD***, 2002

CHASSIOTI, E., WORTHINGTON, D. A new model for call centre queue management. ***Journal of the Operational Research Society***, vol. 55, p. 1352-1357, 2004

CHOKSHI, R. Decision support for call center management using simulation. ***Winter Simulation Conference***, p. 1634-1639, 1999

CHOKSHI, R. Simulation: a key to call center management. ***AT&T Laboratories***, White Paper, 1999

CONTAX. ***Contax Contact Center.*** Disponível em: www.contax.net.br. Acesso em: 19 jul. 2006

CORRAR, L., THEÓPHILO, C. ***Pesquisa Operacional para Decisão em Contabilidade e Administração.*** São Paulo: Atlas, 2004

CRESCE faturamento do setor de telemarketing. ***Estatísticas***. Disponível em: www.callcenter.inf.br. Acesso em: 20 set. 2005

DATABASE SYSTEMS CORP. Products. ***Call Center Simulation Models***. Disponível em: www.databasesystemscorp.com. Acesso em: 20 set. 2005

DECKER, D. Call center's modeling in American Express Company. ***American Express Company***, White Paper, 1999

DIJK, N. Why queuing never vanishes. ***European Journal of Operational Research,*** vol. 99(2), p. 463-476, 1997

ECO, U. ***Como se faz uma tese.*** São Paulo: Perspectiva, 1995

ELANGO, A. Call Center Quality Analysis Using Simulation. ***Asian Simulation Conference***, 2006

EMERALD. ***Advanced Search***. Disponível em: www.emeraldinsight.com. Acesso em: 16 out. 2006

EMPREGO em contact center está em alta. ***Estatísticas***. Disponível em: www.callcenter.inf.br. Acesso em: 20 set. 2005

FREITAS, L. ***Call center: modismo ou realidade?*** Rio de Janeiro: Record, 2000

FUKUNAGA, A., HAMILTON, E., FAMA, J., ANDRE, D., MATAN, O., NOURBAKHSH, I. Staff Scheduling for Inbound Call Centers and Customer Contact Centers. ***American Association for Artificial Intelligence,*** vol. 23(4), p. 30-40, Winter 2002

GANS, N., KOOLE, G., MANDELBAUM, A. Telephone Call Centers: Tutorial, Review and Research Prospects. ***Manufacturing and Service Operations Management***, vol. 5(2), p. 79-141, 2003

GIL, A. ***Métodos e técnicas de pesquisa social.*** São Paulo: Atlas, 1987

GOMES, L. ***O dilema entre qualidade e produtividade no serviço de atendimento ao cliente (SAC): um estudo de caso brasileiro***. Orientador: Kleber Fossati Figueiredo. Rio de Janeiro: UFRJ/COPPEAD, 2004. Dissertação. (Mestrado em Administração de Empresas)

GREEN, L., KOLESAR, P. The pointwise stationary approximation for queues with nonstationary arrivals. ***Management Science***, vol. 37, p. 84-97, 1991

GROSS, D., HARRIS, C. ***Fundamentals of Queueing Theory.*** New York: Willey, 1985

GROSSMAN, T., SAMUELSON, D., OH, S., ROHLEDER, T. Call Centers. ***Encyclopedia of Operations Research and Management Sicence,*** Boston: Kluwer Academic Publishers, p. 73-76, 2001

GULATI, S., MALCOLM, S. Call center scheduling technology evaluation using simulation. ***Winter Simulation Conference,*** 2001

HALL, B., ANTON, J. Optimizing your call center through simulation***. Call Center Solutions Magazine***, Oct. 1998

HANKE, J. ***Business Forecasting.*** New Jersey: Prentice Hall, 1992

HAWKINS, L., MEIER, T., NAINIS, W., JAMES, H. Planning Guidance Document For US Call Centers. ***Information Technology Support Center***. Maryland, 2001

HERTZ, D. Análise de risco em investimentos de capital. ***Biblioteca Harvard de Administração de Empresas***, 8(3), p. 1-14, 1980

HILLIER, F., LIEBERMAN, G. ***Introduction to Operations Research.*** New York: McGraw-Hill, 1995

HILLIER, F., YU, O. ***Queueing Tables and Graphs.*** New York: North-Holland, 1981

IEX. ***TotalView Central***. Disponível em: www.iex.com/TotalView. Acesso em: 20 set. 2005

JENNINGS, O., MANDELBAUM, A., MASSEY, W., WHITT, W. Server Staffing to Meet Time-Varying Demand. ***Management Science***, vol. 42(10), p. 1383-1394, Oct. 1996

KELTON, W., SADOWSKI, R., SADOWSKI, D. **Simulation with Arena**. New York: McGraw-Hill, 2002

KLEINROCK, L. ***Queueing Systems, Volume I: Theory.*** New York: Willey, 1975

KLUNGLE, R. Simulation of a claims call center: a sucess and a failure. ***Winter Simulation Conference***, p. 1648-1653, 1999

KLUNGLE, R., MALUCHNIK, J. The role of simulation in call center management. ***MSUG Conference***, 1997

KOOLE, G., MANDELBAUM, A. Queueing Models of Call Centers: an Introduction***.*** ***Annals of Operations Research,*** 113, p. 41-59, 2002

KOOLE, G., POT, A. Approximate Dynamic Programming in Multi-Skill Call Centers. ***Winter Simulation Conference***, 2005

KUBLER-ROSS, E. Simulação em processos. ***CAUSA & EFEITO – Boletim da Siqueira Campos (Aprendizado instantâneo)***, ano VII(12), 2003

LAM, K., LAU, R. A simulation approach to restructuring call centers. ***Business Process Managament Journal,*** vol. 10, issue 4, Research paper, 2004

LAPPONI, J. ***Estatística usando Excel.*** Rio de Janeiro: Elsevier Editora, 2005

LEVINE, D., BERENSON, M., STEPHAN, D. ***Estatística: Teoria e Aplicações.*** Rio de Janeiro: LTC Editora, 2000

LIMA, L. Telemarketing em expansão. ***O Dia.*** Rio de Janeiro, 12 de março de 2006, p. 1. Caderno Empregos

MANCINI, L. ***Call Center: estratégia para vencer.*** São Paulo: ASK! Cia Nacional de Call Center, 2001

MANDELBAUM, A. Call centres, research bibliography with abstracts. Technical Report, Technion, ***Israel Institute of Technology***, 2003

MANDELBAUM, A., SAKOV, A., ZELTYN, S. Empirical Analysis of a Call Center. Technical Report, Technion, ***Israel Institute of Technology***, 2001

MARKT, P. Driving Large Call Center Simulations Using OLAP Data Cubes. ***Winter Simulation Conference***, 2005

MATAN, O., NOURBAKHSH, I. Playing the Numbers: Using ACD Statistics for Workforce Management. ***Telemarketing & Call Center Solutions***, Mar 1998

MAZZUCHI, T., WALLACE, R. Analyzing Skill-Based Routing Call Centers Using Discrete-Event Simulation and Design Experiment. ***Winter Simulation Conference***, 2004

MEHROTRA, V. Ringing Up Big Business***. OR/MS Today***, 24(4), p.18-24, Aug 1997

MEHROTRA, V., FAMA, J. Call Center Simulation Modeling: Methods, Challenges and Opportunities. **Winter Simulation Conference**, p. 135-143, 2003

MEHROTRA, V., PROFOZICH, D., BAPAT, V. Simulation: the best way to design your call center. ***Telemarketing & Call Center Solutions***, Nov 1997

MILLER, K., BAPAT, V. Case study: simulation of the call center environment for comparing competing call routing technologies for business case ROI projection. ***Winter Simulation Conference***, p. 1694-1700, 1999

MINGHELLI, G. ***Call center: estudo de casos múltiplos em empresas de telefonia do estado do Rio Grande do Sul***. Orientador: Luís Felipe Schilling. São Leopoldo: Universidade do Vale do Rio dos Sinos, 2002. Monografia (Bacharelado em Informática)

MOURA, M. ***Análise e comparação de sistemas de atendimento (Call Center) em diferentes topologias.*** Orientador: Wagner Luiz Zucchi. São Paulo: INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO DE SÃO PAULO, 2004. Dissertação. (Mestrado profissionalizante em Engenharia de Computação)

NAUMES, W., NAUMES, M. ***The Art & Craft of Case Writing.*** Thousand Oaks: Sage, 1999

NOVASIM. ***ccProphet***. Disponível em: www.novasim.com. Acesso em: 20 set. 2005

OUTSOURCING. ***Ranking***. Disponível em: www.callcenter.inf.br. Acesso em: 20 set. 2005

PARAGON Tecnologia. **Introdução à Simulação com ARENA**, 2002.

PARAGON Tecnologia. **Técnicas Avançadas de Simulação com ARENA**, 2002

PARAGON. ***Simulação de Call Center com Arena Contact Center***. Disponível em: www.paragon.com.br. Acesso em: 20 set. 2005

PAZETO, T. ***Metodologia para o Planejamento de Capacidade de Call Center.*** Orientador: Elizabeth Sueli Specialski. Florianópolis: UFSC, 2000. Dissertação. (Mestrado em Ciências da Computação)

PEREIRA Jr., P. ***Avaliação de políticas de estoque e movimentação de contêineres vazios utilizados no transporte marítimo.*** Orientador: Eduardo Saliby. Rio de Janeiro: UFRJ/COPPEAD, 1999. Dissertação. (Mestrado em Administração de Empresas)

PICHITLAMKEN, J., DESLAURIERS, A., L´ECUYER, P., AVRAMIDIS, A., Modelling and Simulation of a Telephone Call Center. ***Winter Simulation Conference***, p. 1805-1812, 2003

PIDD, M. ***Computer Simulation in Management Science.*** New York: Willey, 1998

PORTAGE COMMUNICATIONS INC. ***SimACD - Computer Simulation Software for Inbound Call Centers***. Disponível em: www.portagecommunications.com. Acesso em: 20 set. 2005

RAGSDALE, C. ***Spreadsheet Modeling and Decision Analysis***. Tennessee: South-Western, 2001

REYNOLDS, P. Forecasting Fundamentals: The Art and Science of Predicting Call Center Workload. ***Technology Marketing Corporation***, White Paper, Feb 17, 2005

RILEY, D. Simulating a Virtual Customer Service Center. ***Winter Simulation Conference***, 2005

SAKAMOTO, K. ***Como implementar um call center.*** São Paulo: Happy Few, 2001

SALIBY, E. ***Repensando a Simulação: a Amostragem Descritiva.*** São Paulo: Atlas, 1989

SALIBY, E. Descriptive Sampling: a Better Approach to Monte Carlo Simulation. ***Journal of the Operational Research Society***, vol. 41(12), p. 1133-1142, 1990

SALIBY, E. Descriptive Sampling: an Improvement over Latin Hypercube Sampling. ***Winter Simulation Conference***, p. 230-233, 1997

SALTZMAN, R., MEHROTRA, V. A Call Center Uses Simulation to Drive Strategic Change***. Interfaces*** 31(3), Part 1 of 2, p. 87-101, May-June 2001

SALTZMAN, R., MEHROTRA, V. A Manager-Friendly Platform for Simulation Modeling and Analysis of Call Center Queueing Systems. ***Winter Simulation Conference***, 2004

SHEN, H., BROWN, L. Nonparametric Regression and Confidence Bands with Lognormal Errors and Application to Bank Call Center Data. ***Department of Statistics and Operations Research (the University of North Carolina)***, Technical Report, 05 – 03, 2005

SHEN, H., HUANG, J. Analysis of call centre arrival data using singular value decomposition. ***Applied Stochastic Models in Business and Industry***, vol. 21, p. 251-263, 2005a

SHEN, H., HUANG, J. Forecasting arrivals to a Telephone Call Center. ***Department of Statistics and Operations Research (the University of North Carolina)***, Technical Report, 05 – 05, 2005b

SILVA, E., SILVA, E., GONÇALVES, V., MUROLO, A. ***Pesquisa Operacional.*** São Paulo: Atlas, 1998

SILVA Jr., L. ***Uma Análise para Utilização de Sistemas Workflow em Centrais de Atendimento.*** Orientador: Maria Beatriz Felgar de Toledo. Campinas: UNICAMP, 2004. Dissertação. (Mestrado profissionalizante em Ciência da Computação)

SILVER, M. ***Estatística para Administração.*** São Paulo: Atlas, 2000

SIMULATE. ***A Simulate trabalha com o Simul8***. Disponível em: www.simulate.com.br. Acesso em: 20 set. 2005

SMAILES, J., McGRANE, A. ***Estatística Aplicada à Administração com Excel.*** São Paulo: Atlas, 2002

STECKLEY, S., HENDERSON, S., MEHROTRA, V. Forecast Errors and Random Arrival Rates. ***Montreal Workshop on Call Centers***, 2006

STECKLEY, S., HENDERSON, S., MEHROTRA, V. Performance Measures for Service Systems with a Random Arrival Rate. ***Winter Simulation Conference***, 2005

TAKAKUWA, S., OKADA, T. Simulation Analysis of Inbound Call Center of a City-gas Company. ***Winter Simulation Conference***, 2005

TANIR, O., BOOTH, R. Call center simulation in Bell Canada. ***Winter Simulation, Conference***, p. 1640-1647, 1999

VERGARA, S. ***Projetos e Relatórios de Pesquisa em Administração*** São Paulo: Atlas, 1997

WALLACE, R., SALTZMAN, R. Comparing Skill-Based Routing Call Center Simulations Using C Programming and Arena Models. ***Winter Simulation Conference***, 2005

WEINBERG, J., BROWN, L., STROUD, J. Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data. ***University of Pennsylvania***, White Paper, Jun 2006

WEISSHUHN, W. ***Análise de custos na formação de preços em uma empresa de Call Centers: um estudo de caso***. Orientador: Marcos Gonçalves Avila. Rio de Janeiro: UFRJ/COPPEAD, 2004. Dissertação. (Mestrado em Administração de Empresas)

WHITT, W. Staffing a Call Center with Uncertain Arrival Rate and Absenteeism. ***Production and Operations Management***, vol. 15(1), p. 88-102, Spring 2006

WISE, R. The Visible Difference: Making Informed Decisions with Workforce Management. ***IEX Corporation***, White Paper, 2005

WORTHINGTON, D., WALL, A. Using the discrete time modelling approach to evaluate the time-dependent behaviour of queueing systems. ***Journal of the Operational Research Society***, vol. 50, p. 777-788, 1999

YIN, R. ***Case study research: design and methods.*** Thousand Oaks: Sage, 1994

ANEXO I: *Checklist* da equipe de planejamento para auxiliar a elaboração da previsão de demanda

Este é um documento genérico, ou seja, pode ser utilizado para diferentes tipos de atendimento. Com isso, algumas premissas podem ser incluídas em relação aos questionamentos abaixo visando deixá-lo mais adequado às características de cada produto.

1. Campanhas de Mídia:
Qual a descrição da Campanha, que características possui, qual o Público Alvo?
Qual a data de Lançamento?
Haverá inserções de Mídia?
Se sim qual a grade Mídia?
Qual o impacto esperado (Aumento de Chamadas)?
Haverá impacto no TMA?

2. Introdução de Novos Serviços:
Qual a descrição do NOVO SERVIÇO (Plano), que características possui, qual o Público Alvo?
Qual a data de Lançamento?
Haverá inserções de Mídia?
Se sim qual a grade de Mídia?
Qual o impacto esperado (Aumento de Chamadas)?
Haverá impacto no TMA?

3. Previsão de Base de Assinantes Ativos:
Qual a previsão de crescimento?
Se sim, qual o fator motivador?
Qual a base prevista?

4. Ações de manutenção ou modernização da rede (exemplo: Obras de Expansão):
Qual o cronograma de manutenção/modernização?
Qual a quantidade de clientes será afetada?
Haverá instabilidade dos serviços?
Haverá impacto no TMA?

5. Interpretação dos impactos externos na previsão de chamadas:
Qual a descrição do Impacto?
Qual calltype/Público-Alvo é afetado pelo Impacto?
Haverá impacto no TMA?
Se sim, qual o aumento/redução no volume dos calltypes afetados?

6. Outros eventos que possam influenciar o TMA:
Quais suas características, previsão de início, Calltypes atingidos, e impacto no TMA.

7. Outros eventos que possam influenciar o volume de chamadas:
Quais os eventos, com suas características, previsão de início, Base de Clientes atingida.

Fonte: Material coletado nas entrevistas