

Rogério de Souza Versage

**METAMODELO PARA ESTIMAR A CARGA TÉRMICA DE
EDIFICAÇÕES CONDICIONADAS ARTIFICIALMENTE**

Tese submetida ao Programa de Pós
Graduação em Engenharia Civil da
Universidade Federal de Santa Catarina
para a obtenção do Grau de Doutor em
Engenharia Civil.

Orientador: Prof. Roberto Lamberts, PhD.

Florianópolis
2015

Ficha de identificação da obra elaborada pelo autor através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Versage, Rogério

METAMODELO PARA ESTIMAR A CARGA TÉRMICA
DE EDIFICAÇÕES CONDICIONADAS
ARTIFICIALMENTE / Rogério Versage;
orientador, Roberto Lamberts -
Florianópolis, SC, 2015.

191 p.

Tese (doutorado) - Universidade Federal
de Santa Catarina, Centro Tecnológico.
Programa de Pós-Graduação em Engenharia
Civil.

Inclui referências

1. Engenharia Civil. 2. Desempenho
térmico de edificações. 3. Simulação
computacional. 4. Metamodelos. I. Lamberts,
Roberto. II. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em
Engenharia Civil. III. Título.

Rogério de Souza Versage

**METAMODELO PARA ESTIMAR A CARGA TÉRMICA DE
EDIFICAÇÕES CONDICIONADAS ARTIFICIALMENTE**

Tese julgada adequada para a obtenção do Título de DOUTOR em Engenharia Civil, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Civil - PPGEC da Universidade Federal de Santa Catarina - UFSC.

Prof. Roberto Caldas de Andrade Pinto, PhD. - Coordenador do PPGEC

Prof. Roberto Lamberts, PhD. - Orientador

COMISSÃO EXAMINADORA:

Prof. Fernando Oscar Ruttkay Pereira, PhD. – PPGEC/UFSC

Prof. EneDir Ghisi, PhD. – PPGEC/UFSC

Prof. Dr. Nathan Mendes – PUCPR

Prof. Aldomar Pedrini, PhD. – UFRN

Prof. Dr. Martin Gabriel Ordenes Mizgier – UFSC

AGRADECIMENTOS

Agradeço ao universo pela oportunidade de enfrentar o desafio de um doutorado, desenvolvendo habilidades e consciência para difundir as verdades aprendidas na colaboração ao advento de uma sociedade melhor e mais esclarecida.

Agradeço aos meus pais pelo carinho, apoio e principalmente pela generosidade ao me proporcionar condições para meu aprendizado.

Agradeço ao professor Roberto Lamberts pela sábia orientação. Obrigado por me conduzir nesta caminhada, pelos preciosos conselhos e animações quando estive sujeito a erros e desfalecimentos.

Agradeço ao professor Wagner Augusto Andreasi lembrando as palavras de Frank Shermard Land, "O início é o mais importante". Você desempenhou um papel extraordinário na minha iniciação à ciência e à pesquisa, com ensinamentos que sempre levarei comigo.

Agradeço aos professores, que gentilmente, aceitaram avaliar este trabalho, dedicando seu tempo e sabedoria para contribuir com o desenvolvimento desta tese. Obrigado professores Fernando Oscar Ruttkay Pereira, Eneid Ghisi, Nathan Mendes, Aldomar Pedrini e Martin Gabriel Ordenes Mizgier.

Também agradeço à direção, aos professores e aos funcionários do Programa de Pós Graduação em Engenharia Civil da Universidade Federal de Santa Catarina.

Ainda entre as instituições, agradeço à Eletrobrás, pela bolsa de estudo concedida e pelo financiamento deste trabalho ligado ao Regulamento Técnico da Qualidade do Nível de Eficiência Energética de Edifícios Comerciais, de Serviços e Públicos. E à CAPES, pela bolsa de estudo concedida por alguns meses durante este doutorado.

Agradeço ao Geof Sawaya, do *Lawrence Berkeley National Laboratory*, que através do diálogo estratégico de energia entre Brasil e Estados Unidos colaborou diretamente, possibilitando o desenvolvimento dos sistemas para as simulações utilizadas neste trabalho e trabalhos futuros do LabEEE. E à equipe do SETIC/UFSC pela colaboração na disponibilidade de ferramentas computacionais indispensáveis para o desenvolvimento desta tese.

De forma coletiva, preciso registrar meus agradecimentos aos colegas e amigos do LabEEE, que conviveram comigo os prazeres e dificuldades da vida acadêmica. Obrigado Alexandre Montenegro, Aline Schaefer, Ana Kelly Marinovski, Ana Paula Melo, Andrea Invidiata, Andriago Antonioli, Arthur Santos Silva, Bruna Balvedi, Carolina Carvalho, Clarissa Debiasi Zomer, Cláudia Donald, Cláudia Morishita, Deivis Marinovski, Elisa Beck, Gabriel Iwamoto, Gustavo Palladini, Helena Avis, Joyce Carlo, Juliana Batista, Juliana Cruz, Karran Besen, Laiane Almeida, Leonardo Mazzaferro, Márcio Sorgato, Maria Andrea Triana Montes, Michele Fossati, Miguel Pacheco, Raphaela Walger Fonseca, Renata DeVecchi, Ricardo Rupp, Rosana Debiasi, Solange Goulart, Trajano Viana, Vanessa Duarte, Veridiana Scalco e tantos outros que convivi no laboratório.

Agradeço a meus melhores amigos e irmãos que sempre me incentivaram e me apoiaram a viver distante deles, compreendendo minha ausência em horas importantes de suas vidas.

Agradeço ainda ao querido casal Andrea Triana e Olavo Kucker pela amizade e camaradagem, estes foram peças fundamentais para manter o entusiasmo e enxergar na prática como nosso trabalho pode fazer a diferença na construção de uma humanidade melhor.

Agradeço à família Universitária Alcio Antunes, fraternidade que garantiu o refúgio e fortaleza à minha consciência. Garantindo minha atenção às questões que agitam as sociedades humanas, não me permitindo me fechar ao mundo durante o processo de produção desta tese.

Finalmente, agradeço em especial à presença amorosa, ajuda e motivação de Maiara Dallazen Camillo. Obrigado pela paciência nos momentos mais tensos de trabalho, sua companhia garantiu sanidade e tranquilidade para esta produção.

“Tudo é número.”
Pitágoras

RESUMO

Simulações computacionais são os métodos mais avançados para predição do desempenho energético de edificações. Por outro lado, métodos estatísticos de funções de inferência baseadas em amostras são mais rápidos e simples de serem utilizados. Logo, para combinar as funcionalidades principais de simulações computacionais com o poder das predições estatísticas, existem os modelos híbridos, chamados metamodelos. O objetivo desta tese é desenvolver um metamodelo para estimar a carga integrada anual de energia para refrigeração para avaliação de desempenho energético de edificações condicionadas artificialmente através do desempenho individual de suas zonas térmicas. Foi construída uma base de dados de 1,29 milhões de casos simulados de uma zona térmica, variando parâmetros construtivos e de uso para o clima de Florianópolis. Uma amostra desta base de dados foi utilizada para elaboração de metamodelos com as técnicas de regressão linear múltipla, regressão adaptativa multivariada por splines, processo gaussiano, máquina de vetores de suporte, Randon Forest e redes neurais artificiais. Estes metamodelos foram comparados quanto ao seu desempenho e ao tempo de processamento computacional, e o metamodelo de maior desempenho foi apurado para diferentes tamanhos de amostra e diferentes configurações de sua estrutura. Por fim, um metamodelo de rede neural artificial treinado com aproximadamente 13 mil casos e 72 nós na camada interna reproduziu o resultado de 1,29 milhões de casos com erros menores que 10% para 99,2% dos resultados. Este tipo de metamodelo pode ser aplicados em dimensionamento de sistemas de condicionamento de ar, em tomadas de decisão no desenvolvimento de projetos e na avaliação de desempenho de certificações e regulamentos.

Palavras-chave: Desempenho energético de edificações comerciais; metamodelos; EnergyPlus.

ABSTRACT

Computer simulations are the most advanced methods to predict the energy performance of buildings. Furthermore, statistical methods using inference functions based on samples are faster and simpler to use. Then, to match the main features of computer simulations with the power of statistical predictions, there are developed hybrid models, called metamodels. This thesis aims to develop a metamodel to predict the annual integrated energy cooling load demand for buildings energy performance assessment. A database of 1.29 million thermal zone cases was constructed with a parametric combination of buildings parameters and internal gains. These cases represent thermal zones of commercial buildings to the Florianópolis climate. A sample of this database was used to develop metamodels with different statistical techniques: multiple linear regression; multivariate adaptive regression splines; gaussian process; support vector machine; Random Forest; and artificial neural networks. The metamodels were compared for their performance and computational resources for their development, then the best metamodel was calculated for different sample sizes and different configurations of its structure. Finally, a metamodel of artificial neural network trained with 13 thousand cases and 72 nodes in the inner layer reproduces the result of 1.29 million cases with errors smaller than 10% to 99.2% of results. This type of metamodel can be applied in sizing air conditioning systems, in project development decision-making and evaluation of performance in certifications and regulations.

Keywords: commercial buildings energy performance; metamodels; EnergyPlus.

LISTA DE FIGURAS

Figura 1. Exemplo de gráficos de contorno em um estudo de Screening...	44
Figura 2. Média e desvios padrão do efeito da distribuição das variáveis..	45
Figura 3. Percentual de janela na fachada da zona térmica	74
Figura 4. Representação do AVS e AHS em relação a janela.	76
Figura 5. Representação do AOV em relação a janela.	77
Figura 6. Superfície de obstrução conforme AOV em projeção estereográfica.....	78
Figura 7. Tipos de exposição do piso e cobertura.....	79
Figura 8. Horários de ocupação dos dias de semana.	83
Figura 9. Representação da zona térmica e seus parâmetros variáveis.....	84
Figura 10. Representação do modelo de simulação.....	87
Figura 11. Histogramas de cargas anuais e picos de refrigeração e aquecimento.....	91
Figura 12. Influência dos parâmetros de carga interna na carga de refrigeração anual.	93
Figura 13. Influência dos parâmetros da parede externa na carga de refrigeração anual.	94
Figura 14. Influência dos parâmetros da cobertura na carga de refrigeração anual.	95
Figura 15. Influência dos parâmetros de exposição do piso e cobertura na carga de refrigeração anual.....	96
Figura 16. Influência dos parâmetros da janela na carga de refrigeração anual.	97
Figura 17. Influência dos parâmetros de sombreamento na carga de refrigeração anual.	98
Figura 18. Influência dos parâmetros de orientação e pé-direito da zona térmica na carga de cefrigeração cnual.....	99
Figura 19. Influência do percentual de área de janela por orientação solar na carga de refrigeração anual.....	100
Figura 20. Esquema do método de comparação dos metamodelos.....	104
Figura 21. Matriz de correlação entre as variáveis preditores	105
Figura 22. Distribuição das cargas anuais de refrigeração antes e depois da transformação das unidades.....	108
Figura 23. Esquema de validação cruzada de 3 folds.	113
Figura 24. Tempo de treinamento dos metamodelos	116

Figura 25. Gráficos de dispersão e índices de desempenho dos metamodelos	119
Figura 26. Distribuição de erros reais	122
Figura 27. Esquema do método de refinamento da Rede Neural Artificial.	126
Figura 28. Relação entre o tamanho da amostra e o erro máximo estimado da amostra.....	128
Figura 29. Tempos de treinamento por tamanho de amostra.	138
Figura 30. Tempos de treinamento por tamanho da camada interna.....	139
Figura 31. Coeficiente de determinação de Spearman (R^2) por tamanho de amostra.....	140
Figura 32. Coeficiente de determinação de Spearman (R^2) por tamanho da camada interna.	141
Figura 33. Raiz do erro médio quadrático (RMSE) por tamanho da amostra.	142
Figura 34. Raiz do erro médio quadrático (RMSE) por tamanho da camada interna.	143
Figura 35. Tempo de treinamento por tamanho de amostra e camada interna.	147
Figura 36. Coeficiente de determinação por tamanho de amostra e camada interna.	148
Figura 37. Raiz do erro médio quadrático pelo tamanho da amostra e camada interna.	149
Figura 38. Dispersão dos valores Simulados vs. Preditos dos casos interpolados.....	150
Figura 39. Dispersão dos valores simulados vs. preditos dos casos extrapolados.....	152
Figura 40. Dispersão dos resultados preditos vs. simulados e histograma dos erros.	159
Figura 41. Histograma dos erros relativos	160
Figura 42. Dispersão entre resultados simulados e preditos por faixa de erro relativo.	161
Figura 43. Histogramas da ocorrência de parâmetros em casos com desvios maiores que 5kwh/m ² .ano.....	164

LISTA DE TABELAS

Tabela 1. Análise bibliográfica das limitações do modelo de predição do RTQ-C.....	18
Tabela 2. Tipos de paredes simuladas	71
Tabela 3. Tipos de coberturas simuladas	71
Tabela 4. Parâmetros e valores simulados	85
Tabela 5. Variáveis substitutas	106
Tabela 6. Exemplo de críticos associados ao grau de confiança da amostra	111
Tabela 7. R^2 e RMSE dos treinamentos de validação cruzada	117
Tabela 8. Média e desvio padrão dos resultados simulados e preditos pelos metamodelos	118
Tabela 9. Índices de desempenho global dos metamodelos	123
Tabela 10. Erros máximos estimados para diferentes tamanhos de amostra.	127
Tabela 11. Valores não vistos intermediários e extrapolados.....	132
Tabela 12. Tipos de paredes e coberturas não vistos	133
Tabela 13. R^2 e RMSE dos treinamentos de validação cruzada das Redes Neurais Artificiais	135
Tabela 14. Índices de desempenho das Redes Neurais Artificiais.....	145
Tabela 15. Resultados de R^2 e RMSE dos valores simulados vs. preditos para os casos Interpolados e Extrapolados.	153
Tabela 16. Índices de Desempenho Global com desempenho para casos interpolados.	156

LISTA DE ABREVIATURAS E SIGLAS

aCob	Absortância solar da cobertura
AHL	Amostragem por Hipercubo Latino
AHS	Ângulo Horizontal de Sombreamento
AOV	Ângulo de Obstrução Vizinha
aPar	Absortância solar da parede
AVS	Ângulo Vertical de Sombreamento
CFD	Computational Fluid Dynamics
COBn	Tipo de cobertura n
CTCob	Capacidade Térmica da Cobertura
CTPar	Capacidade Térmica da Parede
DEI	Densidade Equipamentos e Iluminação
DPessoas	Densidade de Pessoas
Exp_Cob	Exposição da cobertura
Exp_Piso	Exposição do piso
FS	Fator Solar
HVAC	Heating, Ventilation and Air Conditioned
IDF	Input data file
IDG	Índice de Desempenho Global
LabEEE	Laboratório de Eficiência Energética em Edificações
MARS	Multivariate adaptive regression splines
MVS	Máquina de vetores de suporte
NRMSE	Normalized Root Mean Square Error
Ocup	Ocupação
PARn	Tipo de parede n
PG	Processo Gaussiano
R²	Coefficiente de determinação
RF	Randon Forest
RLM	Regressão Linear Múltipla
RMSE	Root Mean Square Error
RNA	Rede Neural Artificial
RTQ-C	Regulamento Técnico da Qualidade de Edificações Comerciais, de Serviço e Públicas.
RTQ-R	Regulamento Técnico da Qualidade de Edificações Residenciais.
UCob	Transmitância térmica da cobertura
UFSC	Universidade Federal de Santa Catarina
Upar	Transmitância térmica da parede
UVid	Transmitância térmica da janela
WWR	Window-to-wall Ratio

SUMÁRIO

1.....	Introdução	13
1.1.	Objetivos	21
1.1.1.	Objetivo geral	21
1.1.2.	Objetivos específicos	21
1.2.	Estrutura da tese.....	21
2.....	Revisão Bibliográfica	25
2.1.	Modelos de desempenho termo-energético de edificações	25
2.1.1.	Modelos caixa-branca.....	27
2.1.2.	Modelos caixa-preta	30
2.1.3.	Modelos caixa-cinza	32
2.2.	Desenvolvimento de metamodelos	42
2.2.1.	Pré-processamento de dados.....	42
2.2.2.	Amostragem.....	46
2.2.3.	Métodos de Inferência Estatística	48
2.2.3.1.	Regressão linear múltipla	49
2.2.3.2.	Redes Neurais Artificiais	49
2.2.3.3.	Máquinas de Vetores de Suporte	51
2.2.3.4.	Processo Gaussiano	53
2.2.3.5.	Regressão adaptativa multivariada por splines	53
2.2.3.6.	Random Forest.....	55
2.3.	Aplicação de modelos de edificações	55
2.3.1.	Modelos para cálculo de carga térmica.....	56
2.3.2.	Modelos em Certificações, códigos e regulamentos	57
2.3.3.	Modelos como ferramentas de projetos	60
2.4.	Incertezas nas modelagens	63
2.5.	agilidade em simulações COMPUTACIONAIS.....	64
3.....	Construção da base de dados	69
3.1.	Método: Parâmetros simulados	69
3.1.1.	Parâmetros térmicos	70
3.1.1.1.	Transmitância e Capacidade Térmica.....	70
3.1.1.2.	Absortância solar da parede externa e cobertura.	72
3.1.1.3.	Inércia térmica interna	72
3.1.1.4.	Tipo de vidro	73
3.1.2.	Parâmetros geométricos.....	73
3.1.2.1.	Percentual de janela na fachada.....	74
3.1.2.2.	Sombreamento da abertura	75
3.1.2.3.	Exposição do piso e cobertura	78
3.1.2.4.	Orientação Solar	79
3.1.2.5.	Pé-direito	80

3.1.3.	Carga Térmica Interna	80
3.1.3.1.	Densidade de carga de Iluminação e Equipamentos ...	80
3.1.3.2.	Pessoas	81
3.1.3.3.	Infiltração	82
3.1.3.4.	Horário de ocupação	83
3.1.4.	Total de parâmetros	83
3.2.	Método: Simulações	86
3.2.1.	Zona Térmica	86
3.2.2.	Cluster de simulação	88
3.2.3.	Arquivo Climático	89
3.2.4.	Variáveis dependentes	89
3.3.	Resultados: Apresentação da Base de dados	90
3.4.	Considerações Finais	101
4.	Desenvolvimento e comparação de metamodelos	103
4.1.	Método: Pré-processamento	104
4.1.1.	Exclusão de parâmetros correlacionados	104
4.1.2.	Adição de variáveis substitutas	106
4.1.3.	Transformação da variável dependente	107
4.1.4.	Centralização e escalonamento	108
4.2.	Método: Treinamento dos modelos	109
4.2.1.	Dimensão da amostra para treino	110
4.2.2.	Validação Cruzada	112
4.3.	Método: Avaliação e comparação dos metamodelos	113
4.4.	Resultados: Metamodelos	115
4.5.	Considerações Finais	123
5.	Refinamento do metamodelo de Rede Neural Artificial	125
5.1.	Método: Parâmetros das arquiteturas de rnaS	126
5.1.1.	Tamanho da amostra	126
5.1.2.	Número de nós na camada interna	129
5.1.3.	Avaliação dos modelos	130
5.1.4.	Predição de parâmetros não vistos	130
5.2.	Resultados	134
5.2.1.	Predição de casos não vistos	150
5.2.2.	Análise dos erros da arquitetura da RNA selecionada	158
5.3.	Considerações Finais	166
6.	Conclusões	170
6.1.	Limitações do trabalho	172
6.2.	Sugestões para trabalhos futuros	174
	Referências Bibliográficas	176

1. INTRODUÇÃO

Projetos de edificações são conjuntos de informações formatadas para apresentar conceitos e ideias em forma de dados como dimensões, localização, estética, qualidade, e custo. Estas informações são previsões dos arquitetos e engenheiros do que uma edificação pode vir a ser. São criadas para permitir a comunicação do que cada responsável projetou para a edificação. Hoje em dia, preocupações com o meio ambiente, alterações climáticas, uso racional de recursos, eficiência energética, conforto ambiental e sustentabilidade são considerados de grande importância para o setor de edificações.

Atualmente, a simulação computacional é o método mais avançado para predição do desempenho energético de edificações. As ferramentas de simulação de desempenho de edificações assumem condições dinâmicas de contorno, e normalmente são baseadas em métodos numéricos que buscam aproximações a soluções realísticas considerando as complexidades do mundo real (HENSEM; LAMBERTS, 2011). Entretanto, existem diferentes estratégias para elaboração de modelos de avaliação de desempenho. Há os modelos baseados na resolução de equações físicas que descrevem o comportamento dos sistemas de transferência de calor e os baseados em funções estatísticas para dedução deste comportamento. Os modelos estritamente estatísticos não consideram causas e efeitos, apenas entradas e saídas. Porém, ganham em agilidade dos modelos baseados em equações físicas. Por outro lado, as equações físicas são escritas segundo os princípios da conservação de energia, reproduzindo o comportamento real dos sistemas. Mas para muitos sistemas complexos, o equacionamento dos fenômenos pode ser inviável. Por isso são desenvolvidos modelos híbridos, chamados

metamodelos, que buscam combinar as funcionalidades principais de simulações físicas e o poder de predição estatística.

A utilização de modelos desempenha um importante papel na obtenção de informações de desempenho das edificações. Em arquitetura, maquetes são modelos reduzidos de edificações que possuem algumas características de edifícios reais, que podem ser utilizadas para análises de desempenho ambiental em experimentos em túnel de vento, heliodons e outros equipamentos de análises. Porém, para a análises mais detalhadas de desempenho térmico e energético de edificações, normalmente são utilizados modelos matemáticos que são representações de sistemas que permitem obter conclusões sobre seu desempenho dentro de condições de contorno. Um modelo busca reduzir a complexidade do sistema original sem perder detalhes que sejam importantes para sua descrição. Enquanto a complexidade do próprio modelo é reduzida para simplificar seu desenvolvimento e validação, assim como para agilizar sua aplicação e tempo de análise (AGUIRRE, 2007).

Os modelos de desempenho energético de edificações são utilizados para diversos fins e podem variar sua complexidade de acordo com a aplicação a qual foram desenvolvidos. Em regulamentos, certificações e códigos de energia, os modelos são adotados para predição da eficiência energética da operação de edificações. Etiquetar, dar um selo ou certificar o desempenho energético de edificações são práticas de avaliação que buscam a comunicação de informações da qualidade das edificações aos consumidores. Estes mecanismos visam sensibilizar os consumidores sobre o uso racional de energia e incentivar melhorias na construção através de uma maior transparência no mercado imobiliário.

Historicamente, a eficiência energética esteve negligenciada nos setores de edificações residenciais e comerciais. Restringindo as oportunidades para economia de energia em edificações e limitando a capacidade dos consumidores para comparar o desempenho energético dos edifícios, impedindo-os de utilizar parâmetros de eficiência como fator de escolha. Esta lacuna de informação impede que o mercado imobiliário avalie, de forma precisa, o valor da eficiência energética, limitando a força do mercado que deveria dirigir seus investimentos para edificações eficientes. Globalmente, a classificação e divulgação do nível de eficiência energética de edificações vêm ganhando aceitação com políticas governamentais e de empresas de energia. Mais de 50 países, cidades e estados possuem políticas de avaliação e divulgação em forma de etiquetagem de desempenho. Métodos de avaliação vêm sendo desenvolvidos em diversos países a fim de avaliar e classificar o desempenho de edificações de todo tipo. Tornando a certificação energética de edificações uma tendência mundial (INSTITUTE FOR MARKET TRANSFORMATION, 2015).

No Brasil, a crise de energia de 2001 acarretou na promulgação da lei da Eficiência Energética que definiu uma Política Nacional de Conservação e Uso Racional de Energia. Esta lei determina níveis máximos de consumo de energia, ou mínimos de eficiência energética, de máquinas e aparelhos consumidores de energia fabricados ou comercializados no País, assim como das edificações (BRASIL, 2001). Paralelamente, o PROCEL, programa estratégico do governo criado pelos Ministérios de Minas e Energia e da Indústria e Comércio sob coordenação da Eletrobrás (BRASIL, 1985), lançou em 2003 o PROCEL Edifica, subprograma com foco na eficiência energética das edificações.

Em decorrência, o INMETRO criou, como parte do Programa Brasileiro de Etiquetagem, os regulamentos para determinação do nível de eficiência energética de edificações, com os quais são outorgadas etiquetas que apresentam o desempenho de edificações em relação à energia. A base técnica destes regulamentos foi desenvolvida por convênios entre Eletrobrás/PROCEL Edifica e UFSC, através do LabEEE, e os regulamentos foram discutidos com a sociedade através de representantes dos setores da indústria da construção civil.

Dois regulamentos foram publicados, o primeiro em 2009 com os Requisitos Técnicos da Qualidade para o nível de eficiência energética em edifícios comerciais, de serviço e públicos (RTQ-C), e outro em 2010 para edifícios residenciais (RTQ-R). Ambos têm a possibilidade de avaliação de edificações através de metamodelos de equações de regressão linear múltipla para avaliação da envoltória das edificações (INMETRO, 2009, 2011). A aplicação de modelos em regulamentos deste tipo exige uma precisão satisfatória para garantir o sucesso dos programas de certificação energética de edificações, pois segundo Pérez-Lombard et al. (2009), o sucesso depende da credibilidade da economia de energia real alcançada e no custo-benefício em obter melhores classificações.

Neste sentido, o uso de metamodelos otimiza a predição do desempenho de edificações. Eles compensam sua imprecisão frente a modelos de equacionamento físico pela facilidade de aplicação e velocidade de processamento (MALKAWI; WAEGEL, 2013). Os modelos de simulação com equacionamentos físicos podem ser mal usados devido a sua complexidade, onde as incertezas de especificação de muitas variáveis podem acabar com sua confiabilidade. Além disso, os metamodelos são mais controláveis por tratarem apenas de poucas

variáveis de entrada, o que reduz sua vulnerabilidade nas incertezas de especificação. Ao mesmo tempo, são mais rápidos que simulações de desempenho, pois garantem respostas quase instantâneas, dependendo da técnica estatística adotada.

Os modelos de predição dos sistemas de etiquetagem brasileiros têm sido utilizados em pesquisas acadêmicas para testar a influência de parâmetros construtivos no desempenho energético de edificações e em estudos de caso para apontar limitações de sua aplicação. Dentre as publicações que exploram os modelos de avaliação da envoltória através de análises de variáveis, estudos de casos e comparações do modelo de predição com simulações, foram identificadas as principais críticas e limitações apontadas aos modelos vigentes, apresentados na Tabela 1.

Tabela 1. Análise bibliográfica e limitações do modelo de predição do RTQ-C

Crítica ou Limitação apontada	Citações de limitações e críticas
1. Não diferenciar área envidraçada e proteções solares por orientação solar da fachada	(CASTRO; WESTPHAL, 2011); (FOSSATI; LAMBERTS, 2010)
2. Limitação na representação da volumetria da edificação	(MELO et al., 2011a); (CASTRO; WESTPHAL, 2011); (PEDRINI et al., 2010)
3. Não responder bem à aplicação de vidros de controle solar	(FONTENELLE; OLIVEIRA SANTOS; SANTOS, 2011); (YAMAKAWA; WESTPHAL, 2011); (FOSSATI; LAMBERTS, 2010)
4. Uso de parâmetros ponderados para toda edificação	(FONTENELLE; OLIVEIRA SANTOS; SANTOS, 2011); (AMORIM et al., 2010); (PACHECO et al., 2012)
5. Necessidade de levantamento de dados não significativos na avaliação	(RODRIGUES et al., 2012)
6. Necessidade de especialização técnica profissional para aplicação do método	(RODRIGUES et al., 2012); (GIARETTA; TEIXEIRA; WESTPHAL, 2012)
7. Considerar apenas um tipo de sistema de HVAC (Split)	(CARLO, 2008); (MELO, 2012)
8. Não considerar influências do entorno edificado	(CARLO, 2008)
9. Os pré requisitos de paredes e coberturas podem penalizar uma edificação onde estes itens não são representativos	(CASTRO; WESTPHAL, 2011); (OLIVEIRA; GUTIERREZ, 2012)
10. Impossibilidade das equações descreverem diversas variações volumétricas	(CARLO, 2008)
11. Uso de apenas um padrão de carga térmica interna e padrão de uso e ocupação	(CARLO, 2008); (MELO, 2012)
12. Não considerar influência na iluminação natural e ventilação natural	(CARLO, 2008)

Em resumo, quanto à avaliação da envoltória, a análise bibliográfica apresentada na Tabela 1 demonstra como os modelos do sistema de etiquetagem brasileiro apresentam limitações para representar áreas envidraçadas e proteções solares por orientação solar. Assim como não responde bem à aplicação de vidros de controle solar. Por buscar

representar a edificação completa, os modelos também impossibilitam a avaliação de edificações com volumetrias complexas. Para representar diferenças na envoltória é preciso simplificar os valores com uso de parâmetros ponderados para toda edificação, por exemplo: só é utilizado um valor de ângulo de sombreamento da edificação no modelo, que deve ser a média ponderada de todos os sombreamentos da edificação, independente da orientação solar. Além disso, por avaliar consumo de energia, o modelo foi elaborado com um tipo de sistema de condicionamento de ar, não permitindo a avaliação de sistemas de diferentes eficiências. Também foram apontadas limitações por Carlo (2008), autora do método, como a impossibilidade das equações representarem diversas variações volumétricas, excluindo alguns fatores de forma da abrangência das equações; o uso de apenas um padrão de carga térmica interna e padrão de uso e ocupação; e não considerar influência na iluminação natural e ventilação natural.

Os modelos do RTQ-C compartilham com outros modelos a limitação de não representar diversas variações volumétricas (CATALINA; IORDACHE; CARACALEANU, 2013; CHOU; BUI, 2014; KOROLJA et al., 2013b; LI; HUANG, 2013; MELO et al., 2011a; TSANAS; XIFARA, 2012). Isso ocorre porque, geralmente, o desenvolvimento de modelos é feito para poucas tipologias de edificações inteiras, limitando a abrangência dos modelos aos mesmos tipos volumétricos adotados na sua elaboração (AL GHARABLY; DECAROLIS; RANJITHAN, 2015). Frente as limitações comuns a estes modelos, o foco desta tese é desenvolver um metamodelo explorando um novo paradigma que assume uma avaliação flexível para atender a variedade arquitetônica do setor, adotando o método de zoneamento, onde

cada zona térmica é avaliada individualmente. Este tipo de zoneamento térmico é utilizado em simulações onde não há um sistema de condicionamento de ar definido, recomendado pelo Apêndice G da ANSI/ASHRAE/IESNA Standard 90.1. Por este método, a edificação é zoneada por áreas de mesma densidade de cargas térmicas internas (pessoas, equipamentos e iluminação) e dividida por zonas térmicas interiores e perimetrais. A zona térmica interior é a área localizada a 4,5 metros das fachadas externas e as zonas perimetrais são as áreas entre as fachadas externas e o limite de 4,5 metros adentro, separadas por orientação (ASHRAE, 2013). Desta forma, o método de predição tem o poder de prever o desempenho de qualquer zona térmica da edificação. O desempenho da edificação completa, pavimento ou ambiente será definido pela ponderação das avaliações de suas zonas térmicas.

Metamodelos devem ser elaborados para permitir a visualização dos efeitos das estratégias adotadas com valores de entrada e saída facilmente reconhecidos pelos arquitetos e engenheiros que irão aplicar o método. Portanto, a divisão da avaliação por zonas térmicas permite a avaliação do desempenho de uma edificação considerando as diferenças da envoltória para cada orientação solar. Parâmetros geométricos, térmicos e de carga térmica interna devem ser considerados para compor a base de dados para elaboração de metamodelos de predição. A precisão do metamodelo deve ser resolvida com a adoção de técnicas estatísticas adequadas de amostragem para as simulações e predição.

1.1. OBJETIVOS

1.1.1. OBJETIVO GERAL

O objetivo desta tese é desenvolver um metamodelo para estimar a energia anual integrada da carga térmica para refrigeração de zonas térmicas de edificações condicionadas artificialmente.

1.1.2. OBJETIVOS ESPECÍFICOS

1. Construir uma base de dados de simulações para análises paramétricas e elaboração de modelos estatísticos com variação de parâmetros térmicos, geométricos e de carga térmica interna;
2. Selecionar uma técnica estatística para representar a energia anual integrada da carga térmica de refrigeração para uma zona térmica;
3. Definir a configuração da técnica estatística selecionada para a relação de desempenho e tempo de processamento para elaboração do Metamodelo;
4. Avaliar os erros do metamodelo e identificar os parâmetros preditores mais incertos.

1.2. ESTRUTURA DA TESE

A tese foi estruturada em seis capítulos. Os dois primeiros são a introdução e revisão bibliográfica, os três seguintes apresentam as etapas

metodológicas com seus devidos resultados e o último capítulo com a conclusão da tese.

A introdução, no primeiro capítulo, apresenta o tema deste trabalho e coloca em foco as suas devidas justificativas. Também são apresentados o objetivo geral e específico e esta estrutura da tese.

A revisão bibliográfica, no segundo capítulo, apresenta a investigação sobre a avaliação de desempenho térmico de edificações, esclarecendo as diferenças entre estratégias de modelagem. Ainda são ilustrados como são desenvolvidos metamodelos, suas aplicações e incertezas.

O terceiro capítulo apresenta a construção da base de dados de casos simulados que foi utilizada como universo para as análises estatísticas dos modelos. Amostras da base de dados simulada neste capítulo foram utilizadas nos capítulos seguintes para treinamento e teste dos modelos de predição.

No quarto capítulo foram elaborados seis metamodelos com as técnicas de regressão linear múltipla, regressão adaptativa multivariada por splines, processo gaussiano, máquina de vetores de suporte, Random forest e redes neurais artificiais. Os desempenhos dos metamodelos foram avaliados para predição da base de dados simulada, sendo comparados pelo coeficiente de determinação, raiz do erro médio quadrático e tempo de processamento computacional para elaboração de cada modelo. A partir desta avaliação foi definida qual técnica estatística alcançou maior precisão para a predição.

No quinto capítulo, a técnica estatística para predição com melhor desempenho apresentada no capítulo anterior foi explorada para analisar seu desempenho com diferentes configurações de sua estrutura. Foram

analisados os erros apresentados pelo modelo e foi definida a melhor estrutura para elaboração de metamodelos para predição da energia anual integrada da carga térmica de refrigeração de zonas térmicas de edificações condicionadas artificialmente.

O sexto capítulo fornece as discussões finais sobre as conclusões desta tese, como um resumo das principais conclusões dos capítulos anteriores, descrevendo as limitações do trabalho e sugestões de trabalhos futuros.

2. REVISÃO BIBLIOGRÁFICA

Primeiramente os tipos de modelos de avaliação de desempenho energético de edificações são apresentados. Em seguida são apontados os métodos de desenvolvimento de metamodelos, desde o tratamento da base de dados até a seleção de técnicas de inferência estatística.

Uma revisão sobre as formas de aplicação de modelos de desempenho energético de edificações também é apresentada, com exemplos de aplicação de modelos em cálculos de demanda de energia da carga térmica para refrigeração e aquecimento, certificações de eficiência energética e em ferramentas de auxílio ao desenvolvimento de projetos arquitetônicos e de engenharia.

Por fim, as incertezas na aplicação de modelos e as inovações na avaliação de desempenho energético são apresentadas, brevemente, para demonstrar as preocupações e as inovações na modelagem do desempenho térmico de edificações.

2.1. MODELOS DE DESEMPENHO TERMO-ENERGÉTICO DE EDIFICAÇÕES

Modelos são representações de sistemas que permitem obter conclusões sobre seu desempenho dentro de condições de contorno, quando desenvolvidos apropriadamente. Um modelo busca reduzir a complexidade do sistema original sem perder detalhes que sejam importantes para sua descrição. A complexidade do modelo pode ser reduzida para simplificar seu desenvolvimento e validação, assim como para agilizar sua aplicação e tempo de análise (AGUIRRE, 2007).

Em arquitetura, a maquete é um tipo de modelo bastante utilizado. É um análogo reduzido de uma edificação que possui algumas características do edifício real. Porém, guardadas as proporções de escala e detalhamento. Para uma análises mais detalhadas de desempenho térmico e energético de edificações são utilizados, normalmente, modelos matemáticos que representam as características de uma edificação como um análogo matemático.

Existem diversas formas de modelagem matemática, porém, de forma geral, os modelos matemáticos podem ser classificados como caixa-branca, caixa-preta ou caixa-cinzas. Esta nomenclatura é baseada nos processos de identificação de sistemas da teoria geral dos sistemas.

Modelos do tipo caixa-branca são baseados na física ou natureza do processo a ser representado, ou ainda como um modelo conceitual. Na modelagem de desempenho térmico de edificações, os modelos caixa-branca são aqueles que reproduzem o comportamento térmico das edificações. Esta abordagem é utilizada em diferentes escalas através de equações de transferência de calor e umidade.

Os modelos do tipo caixa-preta são modelos fechados, quando sua estrutura interna é desconhecida. Em desempenho energético de edificações, estes modelos são aqueles que se utilizam de processos estatísticos em que não é considerado o comportamento físico da edificação. São utilizados quando a correlação entre os valores de saída e entrada não são conhecidos ou de complexidade potencialmente alta.

Já os modelos caixa-cinza são uma mescla entre as características dos modelos caixa-branca e caixa-preta. Esta mistura normalmente se dá pela utilização dos parâmetros físicos e geométricos das edificações com processos estatístico na predição de resultados de desempenho. Também

são chamados de modelos híbridos, por terem seu desenvolvimento baseados em elementos estatísticos e de comportamento físico.

Os modelos podem ser classificados de diferentes formas, porém, geralmente considerando o mesmo conceito geral. Artigos de revisões sobre tipos de modelos são apresentados por Fouquier et al. (2013), Fumo (2014) e Zhao e Magoulès (2012)

2.1.1. MODELOS CAIXA-BRANCA

Os modelos caixa-branca são elaborados a partir de conceitos fundamentais, axiomas e postulados. Estes tipos de modelos também são chamados de modelos físicos, fenomenológicos ou conceituais; ou ainda, conhecidos como caixa de vidro, caixa transparente ou caixa aberta. Para elaboração de modelos caixa-branca se faz necessário conhecer a fundo o sistema a ser modelado, estar familiarizado com os fenômenos físicos envolvidos e conhecer as relações matemáticas para descrevê-los.

Modelos caixa-branca para análise de desempenho de edificações normalmente são modelados para representar o comportamento físico de um tipo de sistema específico como, por exemplo: modelos de condução, convecção e radiação para superfícies opacas ou transparentes, modelos higrotérmicos, modelos de ventilação natural, modelos de sistemas de condicionamento de ar e ventilação e modelos de iluminação e acústica. Também podem representar outros sistemas de edificações, como sistemas fotovoltaicos, aquecimento de água, conforto ambiental e modelos de emissão de poluentes. Sempre baseados na resolução de equações que descrevam o fenômeno físico da transferência de calor que envolve o sistema.

Os modelos físicos podem ser modelados com diferentes níveis de complexidade. Podem ser simples como o modelo de Liao e Dexter (2004), que apresentam um modelo físico simplificado para calcular a temperatura média de um sistema de aquecimento de ar destinado a melhorar seu controle; ou extremamente complexos, como simulações de dinâmica de fluidos computacionais (STEPHENS; GORISSEN; DHAENE, 2011). Estes diferentes níveis de complexidade permitem diferentes níveis de análises. Por exemplo, um modelo caixa-branca permite saídas detalhadas em diferentes escalas de tempo, com valores anuais, mensais, horários e até por minutos; em diferentes escalas espaciais, com resultados para uma edificação completa, uma zona térmica ou uma porção de ar em um ambiente.

Existem vários programas computacionais para resolução de modelos físicos de desempenho térmico e energético de edificações. São chamados de programas de simulação computacional de desempenho termo-energético aqueles que integram diferentes modelos físicos para processar a interação de diversos sistemas da edificação. A vantagem do uso de programas computacionais está no processamento de um grande volume de dados em modelos físicos complexos (CLARKE, 2001). Uma revisão com vários programas deste tipo é apresentada em detalhe por Crawley et al. (2005), que apesar de não ser atual, apresentam uma análise de 20 programas de simulação energética que ainda são referência hoje em dia.

Diversos programas de simulação computacional de desempenho energético de edificações estão disponíveis. Crawley et al. (2005) faz um sumário de alguns deles, onde pode-se destacar o ESP-R, TRNSYS, eQuest, IES, TRACE, PowerDomus e EnergyPlus, que fazem simulações

computacionais integrando vários sistemas das edificações. Segundo Cóstola (2011), os programas geralmente são focados para domínios específicos de geometria, combinados com um ou mais modelos físicos. Entre os programas para simulação de calor, ar e umidade da edificação completa, o autor identifica três tipos de programas: simulação energética de edificações; simulação de energia, ar e umidade de elementos construtivos; e por dinâmica dos fluídos computacionais.

Para este trabalho, o programa de simulação computacional adotado é o EnergyPlus, que permite avaliar o desempenho energético da edificação completa, seu comportamento térmico e outros sistemas. O EnergyPlus é um programa de simulação computacional em constante progresso, seu lançamento foi a continuação de dois programas com histórico de mais de 20 anos de desenvolvimento: o DOE-2, do final dos anos 1960; e o BLAST, do início dos anos 1970. Ambos com a integração de modelos de múltiplos autores e com décadas de investimento e desenvolvimento (CRAWLEY; LAWRIE, 2001; ENERGYPLUS, 2013a). Apesar de largamente utilizado o EnergyPlus possui limitações, onde Cóstola (2011), por exemplo, destaca as limitações em seus modelos de umidade.

Para aplicação de modelos do tipo caixa-branca também existem interfaces gráficas para usuários que simplificam sua complexidade. O EnergyPlus, por exemplo, é utilizado como máquina de simulação da ferramenta S3E – Simulador de Eficiência Energética de Edificações (LOPES et al., 2011). Esta ferramenta, através de uma interface web simples, permite a simulação completa de uma edificação com o EnergyPlus para avaliações pelo método de simulação do sistema de etiquetagem de eficiência energética de edificações brasileiro.

2.1.2. MODELOS CAIXA-PRETA

Diferente dos modelos caixa-branca, os caixa-preta não utilizam equações de transferência de calor, parâmetros físicos ou geométricos. Estes modelos utilizam funções estatísticas baseadas em amostras para dedução do comportamento de sistemas. Sua principal vantagem é dispensar o entendimento dos fenômenos físicos complexos ou desconhecidos, envolvidos no sistema.

Modelos caixa-preta não têm o compromisso de relacionar parâmetros físicos com o desempenho energético de uma edificação. Podendo, por exemplo, correlacionar o consumo de eletricidade com indicadores socioeconômicos, como apresentado por Kuyuk et al. (2010) e Kavaklioglu (2011). Ambos utilizaram modelos estatísticos para estimar o consumo elétrico da Turquia a partir do histórico dos dados de população, produto interno bruto, e importação e exportação de mercadorias. O primeiro utilizou a técnica de Máquina de Vetores de Suporte com dados de 31 anos, e o segundo, Redes Neurais Artificiais com dados de 24 anos. Os dois conseguiram altos coeficientes de determinação na predição do consumo elétrico do país. É comum, em modelos de caixa-preta, se utilizar de dados históricos, pois a relação entre parâmetros de entrada e saída não são conhecidos. Karampelas et al. (2010) e Ekonomou (2010) fizeram o mesmo para Grécia, com dados de 15 anos. Porém, ao invés de utilizarem o ano como parâmetro de entrada, consideraram a temperatura anual média do ar.

Datta, Tassou e Marriott (1997), Karatasou, Santamouris e Geros (2006) e Li et al. (2009), apresentam modelos caixa-preta que utilizam dados históricos de temperatura de bulbo seco, umidade relativa e radiação solar como parâmetros de entrada, e energia anual integrada da

carga térmica de refrigeração como saída para o treinamento de Redes Neurais Artificiais. Com seus modelos, conseguiram prever a carga térmica dos edifícios estudados a partir de dados climáticos horários. Estes tipos de modelos utilizam parâmetros físicos do ambiente externo, porém desconhecem o comportamento térmico da edificação. Como em cada um destes modelos não se variou a edificação, o comportamento do edifício acaba sendo uma variável isolada, permitindo que alcançassem correlações satisfatórias para seus autores.

A modelagem caixa-preta, apesar de não exigir conhecimento do comportamento físico dos sistemas que analisa, requer conhecimento dos processos estatísticos para se trabalhar com análise de dados. Kuhn e Johnson (2013) publicaram o livro *Applied Predictive Modeling* que enfoca o processo de desenvolvimento de modelos de predição baseados, principalmente, em mineração de dados, reconhecimento de padrões e aprendizado de máquina. Os autores apresentam todo o processo de modelagem utilizando a linguagem R (R, 2014), uma linguagem de programação em código aberto para processamentos estatísticos e gráficos. Esta publicação dá suporte para utilização do pacote ‘caret’ (*Classification And Regression Training*) desenvolvido para R (KUHN, 2015). Este pacote disponibiliza funções para gerenciar o desenvolvimento de modelos com várias técnicas de predição. O pacote possui ferramentas para tratamento de dados, pré-processamento de variáveis, seleção de configurações de modelos e avaliação da importância de variáveis.

Os modelos caixa-preta são estritamente estatísticos, desconsiderando os efeitos físicos do comportamento dos sistemas. As análises de modelos caixa-preta não consideram causas e efeitos, apenas

entradas e saídas. Porém, estes modelos ganham em agilidade dos modelos caixa-branca, além de prever sistemas complexos ou com variáveis de natureza desconhecida com mais praticidade. Por sua vez, as equações dos modelos caixa-branca são escritas segundo os princípios da conservação de energia, garantindo o comportamento real dos sistemas. Porém, para muitos sistemas complexos, o equacionamento dos fenômenos pode ser inviável. Por isso, são desenvolvidos modelos híbridos, chamados caixa-cinza, metamodelos ou *surrogate models*.

2.1.3. MODELOS CAIXA-CINZA

Os modelos caixa-cinza buscam combinar as funcionalidades principais de simulações físicas dos modelos caixa-branca e o poder de predição estatística dos modelos caixa-preta. Estes modelos também são chamados de modelos híbridos, por ter seu desenvolvimento com elementos estatísticos e de comportamento físico. A porção de física e estatística pode variar na elaboração dos modelos, podendo-se dizer que existem modelos de diferentes tons de cinza.

Foucquier et al. (2013) relacionam três estratégias de combinação de modelos físicos e estatísticos que resultam em modelos caixa-cinza. A primeira consiste em usar o aprendizado de máquina para estimar parâmetros físicos, normalmente utilizados em tempo real para operação de sistemas. A segunda estratégia é usar um modelo estatístico, treinado a partir de resultados de simulações computacionais de modelos físicos, para prever o desempenho de uma edificação. E a terceira estratégia consiste em usar métodos estatísticos em campos de incertezas dos modelos físicos, em estudos de sensibilidade, por exemplo.

A segunda estratégia relacionada anteriormente é o foco de abordagem deste trabalho. Quando em avaliações de desempenho energético de edificações os modelos caixa-cinza aparecem como métodos simplificados. Estes métodos são baseados em regressões montadas a partir de resultados de simulações termo-energéticas e utilizados para avaliar o desempenho de edificações ou elementos construtivos. Em engenharia, este tipo de modelo caixa-cinza é explorado pelos chamados metamodelos ou *Surrogate Models*.

Metamodelos ou *Surrogate Models* simplificam a obtenção de resultados, simulando o comportamento de um universo completo a partir de uma amostra. Esta avaliação é feita através do comportamento das entradas e saídas de um sistema, independente do processo para se chegar ao resultado. Os metamodelos são construídos a partir dos resultados de simulações de um número limitado de casos, escolhidos de forma sistemática, que alimentam um modelo estatístico para prever o resultados de todos os casos do universo desejado (FORRESTER; SOBESTER; KEANE, 2008; QIAN et al., 2006)

O desenvolvimento deste tipo de modelo não é recente, Crawley et al. (1993) desenvolveram um modelo simplificado do tipo caixa-cinza para o código de eficiência energética do Canadá. Como modelo físico, os autores simularam cargas de refrigeração e aquecimento de 5.400 casos processados pelo programa DOE-2 para 25 cidades canadenses. Foram variados valores de parâmetros construtivos e geométricos de área e transmitância térmica de paredes e janelas, coeficientes de sombreamento e cargas térmicas internas de pessoas, equipamentos e iluminação. Como modelos estatísticos foram desenvolvidas correlações para predição das cargas de refrigeração e aquecimento a partir dos resultados das

simulações. Este tipo de modelo se baseia muito mais nos aspectos físicos do que nos estatísticos, quando os autores buscam interpretar os fenômenos e correlações de cada parâmetro com a variável dependente.

Os dados físicos dos modelos podem ser obtidos de diferentes fontes, Aydinalp, Ugursal e Fung (2002) desenvolveram um modelo de predição de consumo de energia de edificações residenciais a partir de dados de levantamento detalhado de 8.767 residências e dados de consumo de eletricidade de um ano completo de 2.050 residências. Seu modelo caixa-cinza utilizou a técnica de Redes Neurais Artificiais para prever o consumo de equipamentos, iluminação e refrigeração e obteve um coeficiente de determinação de 0,909.

Esse tipo de modelo também pode ter função na otimização de controle de sistemas de condicionamento de ar, como apresenta Zhou et al. (2008), com um modelo caixa-cinza para previsão da carga térmica para o dia seguinte baseado em modelos de previsão do tempo integrado a um modelo de predição de carga térmica de edificações.

Como um exemplo da terceira estratégia de tipo de caixa-cinza, usando métodos estatísticos em campos de incertezas dos modelos físicos, De Wilde, Tian e Augenbroe (2011) propuseram uma metodologia para predição do desempenho de edifícios durante seu ciclo de vida, considerando a degradação dos elementos construtivos e as mudanças climáticas. O método utiliza modelos estatísticos determinísticos e estocásticos para determinar os efeitos do tempo nas condições climáticas e nas propriedades dos materiais. Os autores ilustraram a utilização do método através de simulações paramétricas com EnergyPlus e arquivos climáticos estatisticamente alterados para representar até 40 anos futuros.

Como um metamodelo, Stephens, Gorissen e Dhaene (2011) desenvolveram um método para substituição de simulações em CFD (*Computational Fluid Dynamics*), que são caras e demoradas. Os autores utilizaram a ferramenta SUMO (*Surogate Model toolbox*) para Matlab e estudaram a sensibilidade em fluxos de fluidos em tubulações, comparando o desempenho de modelos de Rede Neural Artificial e Máquina de Vetores de Suporte. Do mesmo modo, Qin et al. (2012), desenvolveram Redes Neurais Artificiais para substituição de análises de CFD para estudos de fluxo de ar em átrios de edificações comerciais.

Em análise térmica de edificações, a porção física do modelo caixa-cinza adota os parâmetros de interesse da análise. Para simplificar a aplicação das funções estatísticas, alguns pesquisadores buscam reduzir ou simplificar estes parâmetros, como Catalina; Iordache e Caracaleanu (2013) que adotaram três parâmetros de entrada e um de saída para elaboração de uma regressão linear múltipla. Foram utilizados resultados de medição e valores de Coeficiente de perda de calor global, temperatura sol-ar, índice de fachada Sul equivalente e as temperaturas de termostato do sistema de condicionamento de ar, alcançando um coeficiente de correlação 0,987 para a predição de consumo de um edifício comercial.

Korolija et al. (2013b) alcançaram coeficientes de correlação entre 0,95 e 0,98 na predição das cargas de refrigeração e aquecimento de edifício de escritórios. Um modelo de regressão linear múltipla foi utilizado para representar o estoque edificado a partir da variação de forma, materiais, envidraçamento e orientação a partir do resultado de 23.040 casos simulados com o programa EnergyPlus para o clima de Londres.

Para o desenvolvimento de metamodelos, diversas pesquisas buscam comparar o desempenho das técnicas estatísticas para predição dos resultados de modelos físicos. Li e Huang (2013) compararam as técnicas de ARMAX (*Autoregressive Moving Average with Exogenous inputs*), Regressão Linear Múltipla, Rede Neural Artificial, e Rede Resistor-Capacitor para predição do desempenho de um edifício de escritórios simulados com o programa TRNSYS. Foram utilizados resultados de 60 simulações variando os parâmetros de massa térmica interna e externa e valores de termostato do sistema de condicionamento de ar.

Fonseca, Didoné e Pereira (2013) desenvolveram metamodelos para avaliação do impacto da iluminação natural no consumo de energia de edificações comerciais com o uso de Redes Neurais Artificiais em comparação a Regressões Lineares Múltiplas. Os autores utilizaram resultados de 216 simulações integradas do programa Daysim com o EnergyPlus e demonstraram que os metamodelos de Redes Neurais artificiais apresentam melhores performances, com coeficiente de determinação de até 0,9867, contra 0,8028 alcançado pelos metamodelos de regressão linear múltipla.

Chou e Bui (2014) compararam o desempenho de várias técnicas de modelagem estatística para a predição das cargas de refrigeração e aquecimento de edificações. As técnicas adotadas foram máquina de vetores de suporte, redes neurais artificiais, árvore de regressão e classificação, interação automática chi-quadrado e regressão linear múltipla. Os modelos de melhor desempenho foram combinados por um algoritmo específico para produzir uma função única. Sendo que a predição do modelo combinado é obtida pela combinação linear das

funções de cada modelo. Segundo os autores, esta combinação pode ser construída facilmente e a expectativa é de sempre se obter um modelo combinado de melhor desempenho que os modelos individuais. Chou e Bui (2014) construíram seus modelos de predição a partir de uma base de dados de Tsanas e Xifara (2012), com 768 simulações energéticas feitas com o programa Ecotect. Esta base de dados é composta por 12 tipos de edificações de diferentes formatos, mas com o mesmo volume interno e materiais construtivos. Os parâmetros avaliados foram as variações de dimensão das paredes, cobertura e janela, num total de oito variáveis independentes. Como variáveis dependentes foram utilizados os resultados simulados de cargas de refrigeração e aquecimento.

Como critérios de comparação entre os modelos, Chou e Bui (2014) adotaram o coeficiente de determinação (R^2), a raiz do erro médio quadrático (RMSE), o erro absoluto médio (MAE) e erro percentual médio absoluto (MAPE). O RMSE é calculado pela raiz quadrada da soma dos quadrados dos erros entre cada resultado predito e seu valor verdadeiro. Em contraste com o RMSE, o MAE é a medida de quão perto as predições estão dos valores reais. O MAE é calculado pela média dos erros absolutos da predição e o MAPE é o valor percentual destes erros. Estes critérios de desempenho foram combinados em um índice de síntese (SI), como uma média da soma das razões do desempenho de cada modelo em comparação com os modelos de maior e menor desempenho. O SI é calculado por meio da Equação 1.

$$SI = \frac{1}{n} \sum_{i=1}^n \left(\frac{P_i - P_{min,i}}{P_{max,i} - P_{min,i}} \right) \quad (1)$$

Onde:

n é o número de critérios de desempenho;

P_i é o indicador de desempenho.

O valor de SI varia de zero a um, onde valores mais próximo a zero indicam alta precisão do modelo de predição. Como resultados, os valores do índice de síntese (SI) variaram entre 0,11 e 0,82 para refrigeração e entre 0,00 e 1,00 para aquecimento. Os melhores resultados foram do modelo combinado de máquina de vetores de suporte com redes neurais artificiais e do modelo de máquina de vetores de suporte. Estes modelos apresentaram erro percentual absoluto médio (MAE) abaixo de 4%, além dos outros critérios de desempenho considerados satisfatórios. Além da avaliação do desempenho de cada modelo, Chou e Bui (2014) avaliaram o tempo de execução das predições, para avaliar o tempo de resposta de sua aplicação. Estes tempos variaram entre aproximadamente meio e um segundo.

O trabalho de Chou e Bui (2014) se destaca pelo desenvolvimento estatístico de modelos caixa-cinza. Entretanto, o uso de resultados de simulações com o programa Ecotect abate sua confiabilidade, pois este programa de simulação não é aplicável à validação da Ashrae Standard 140, conhecida como Bestest (ASHRAE, 2004). Não sendo recomendado para simulações térmicas de edificações, conforme o trabalho de validação apresentado por Vangimalla et al. (2011).

No Brasil, mais especificamente no LabEEE (Laboratório de Eficiência Energética em Edificações), uma sequência de trabalhos desenvolveram modelos do tipo caixa-cinza para predição do

desempenho térmico de edificações. Nos últimos quinze anos os modelos desenvolvidos no laboratório vêm evoluindo tanto na abordagem física, quanto na estatística. Estes modelos vêm sendo aplicados em importantes iniciativas de eficiência energética de edificações no país, como nos programas brasileiros de etiquetagem de eficiência energética de edificações residenciais e comerciais, de serviço e públicas. A seguir são apresentados os principais trabalhos de modelagem desenvolvidos pelo Laboratório.

Iniciando com Signor (1999), foram desenvolvidas equações de modelos para predição de consumo de energia elétrica de edifícios comerciais artificialmente condicionados. Estas equações foram elaboradas a partir de resultados de simulações computacionais com o programa VisualDOE, variando parâmetros construtivos de número de pavimentos, forma, transmitância e absortância das fachadas e cobertura, percentual de área envidraçada das fachadas, coeficientes de sombreamento das janelas e densidade de carga térmica interna. No total, foram simulados 1.024 casos divididos entre climas de 14 capitais brasileiras. Como modelos estatísticos, foram elaboradas equações de regressão linear múltipla para 14 cidades, onde para maioria delas alcançou-se um coeficiente de correlação de 0,99 (SIGNOR; WESTPHAL; LAMBERTS, 2001; SIGNOR, 1999).

Na sequência, agora simulando com o programa EnergyPlus (versão 1.2.3), Westphal e Lamberts (2007) desenvolveram equações de regressão linear múltipla para predizer o consumo de energia de edificações comerciais no Brasil. Foram variados valores de 17 parâmetros físicos, duas tipologias de edificações comerciais e três cidades (Curitiba, Salvador e Florianópolis). Foram simulados 23.040

casos, distribuídos entre cada tipo de edificação e cidade. Por fim, através de amostragem por Hipercubo Latino, reduziram o número de simulações e aumentaram a variação de parâmetros, alcançando um coeficiente de determinação de 0,973.

Seguindo a mesma linha de desenvolvimento, Carlo e Lamberts (2008) lançaram um modelo simplificado, chamado de método prescritivo para avaliação de envoltórias de edificações do programa de etiquetagem de eficiência energética de edifícios comerciais brasileiros. A partir do método desenvolvido por Carlo (2008), os metamodelos de regressão linear múltipla calculam indicadores de consumo de eletricidade a partir de parâmetros construtivos da edificação, como percentual de área envidraçada, fator solar dos vidros, ângulos de proteção solar das janelas, transmitância térmica das paredes e cobertura e áreas e volume da edificação. Foram simulados 5.000 casos de modelos físicos com o programa EnergyPlus, utilizados para elaborar diferentes equações de regressão, divididas por grupos e climas. Os autores alcançaram equações com coeficientes de determinação de 0,9824. Estas equações, reproduzidas para seis climas brasileiros, compuseram o método prescritivo de avaliação da envoltória adotado pelo Programa Brasileiro de Etiquetagem para edificações comerciais, de serviço e públicas (INMETRO, 2009).

Para o programa de etiquetagem de eficiência energética de edificações residenciais foram elaborados modelos do mesmo tipo. Porém, para edificações residenciais, a envoltória é avaliada quanto ao desempenho para frio e calor quando naturalmente ventilada, e para consumo de refrigeração quando condicionada artificialmente. Os modelos foram gerados a partir de resultados de simulações

computacionais com EnergyPlus variando parametricamente valores geométricos e físicos dos elementos construtivos da edificação. Um diferencial deste modelo é que cada ambiente de permanência prolongada da residência é avaliado individualmente pelo metamodelo. No total foram utilizados resultado simulados de 187.200 ambientes distribuídos na elaboração de 20 equações para 7 climas brasileiros. Os coeficientes de determinação alcançados variaram entre 0,42 e 0,91 (VERSAGE; LAMBERTS, 2011). O menor valor de coeficiente de determinação foi para a predição de desempenho para calor nos climas frios da Zona Bioclimática 1, onde a baixa correlação é resultado da baixa demanda por refrigeração. Descartando esta equação, as outras 19 foram adotadas para o método prescritivo do Programa Brasileiro de Etiquetagem para edificações residenciais (INMETRO, 2011).

Recentemente, buscando melhorias na precisão das equações de avaliação de edificações comerciais, Melo (2012) desenvolveu um novo método para estimar o consumo de energia de edificações comerciais através da aplicação de redes neurais artificiais, do qual avanços significativos no método de predição do regulamento foram discutidos. A autora avaliou a precisão do modelo atual do regulamento e explorou comparações entre simulações paramétricas e amostragem por hipercubo latino para o desenvolvimento de redes neurais artificiais como proposta de um novo método simplificado para o Programa Brasileiro de Etiquetagem de edifícios comerciais, de serviço e públicos (MELO et al., 2012, 2011; MELO; LAMBERTS, 2013; MELO, 2012).

2.2. DESENVOLVIMENTO DE METAMODELOS

Os metamodelos devem ser elaborados a fim de reproduzir sistema, utilizando o mínimo de avaliações de simulações. O processo compreende três principais etapas: *Screening*, experimento preliminar para seleção de parâmetros e valores relevantes; amostragem de um conjunto representativo; e modelagem de inferência, treinamento do modelo estatístico. Quando a natureza da verdadeira função da predição desejada não é conhecida, não há como definir qual modelo de predição será o mais preciso para o metamodelo. Os modelos de predição mais utilizados em metamodelos são: superfícies de resposta, processo gaussiano (Krigagem), máquinas de vetores de suporte e redes neurais artificiais (GORISSEN; COUCKUYT, 2010)

2.2.1. PRÉ-PROCESSAMENTO DE DADOS

Kuhn e Johnson (2013) destacam os principais passos no desenvolvimento de um modelo de predição. A importância do tratamento dos dados antes da modelagem depende da técnica de predição a ser utilizada. Modelos de regressão linear, por exemplo, são sensíveis à forma como os dados preditores são apresentados. Já os modelos baseados em árvore de decisões não requerem muito tratamento de dados. Geralmente, as técnicas de pré-processamento de dados buscam adicionar, excluir ou transformar suas variáveis. A transformação de valores reduz o impacto de dados com distribuição enviesada ou dados espúrios. A exclusão de variáveis com alta correlação entre si também é importante, melhorando o desempenho do modelo. Já a adição de variáveis se faz a partir da utilização de variáveis substitutas, resultantes de uma função dos dados

existentes. Novas variáveis podem ser geradas através da multiplicação e divisão dos valores de outras variáveis, assim como na separação de variáveis fatoriais em binários. Segundo os autores, estratégias simples no tratamento dos dados podem melhorar significativamente o desempenho dos modelos.

Minimizar o número de variáveis é um passo importante para atender os objetivos de um Metamodelo da maneira mais eficiente possível. Os procedimentos para reduzir o número de parâmetros são conhecidos como *screening* – peneirada ou peneiração. Porém, esta redução não pode comprometer a relevância das análises. O *screening* deve ser feito quando há a desconfiança de que alguma variável pode ser descartada do estudo. Os algoritmos de Morris são indicados para o *screening* de pré-processamento de dados no desenvolvimento de metamodelos (FORRESTER; SOBESTER; KEANE, 2008)

O método de Morris tem como objetivo estimar os parâmetros da distribuição dos efeitos elementares associados a cada variável. O princípio deve ser que uma grande medida de tendência central indica a variável com influência mais importante sobre a função, e uma grande medida de propagação indica uma variável envolvida em interações e/ou envolvida em uma função não linear. Segundo Forrester, Sóbester e Kean (2008), para um experimento preliminar de *Screening* devem ser definidas as variáveis independentes com valores máximos, mínimos e de base. Os valores do intervalo entre máximo e mínimo são utilizados para gerar gráficos de contorno da função de resultados de pares de variáveis independentes, mantendo as outras variáveis com o valor base. A Figura 1 apresenta um exemplo de gráficos de contorno gerados para um *screening* de parâmetros envolvidos no peso de asas de avião. Cada

parâmetro em X é combinado com os mesmos parâmetros em Y, havendo a combinação de cada parâmetro com todos os outros.

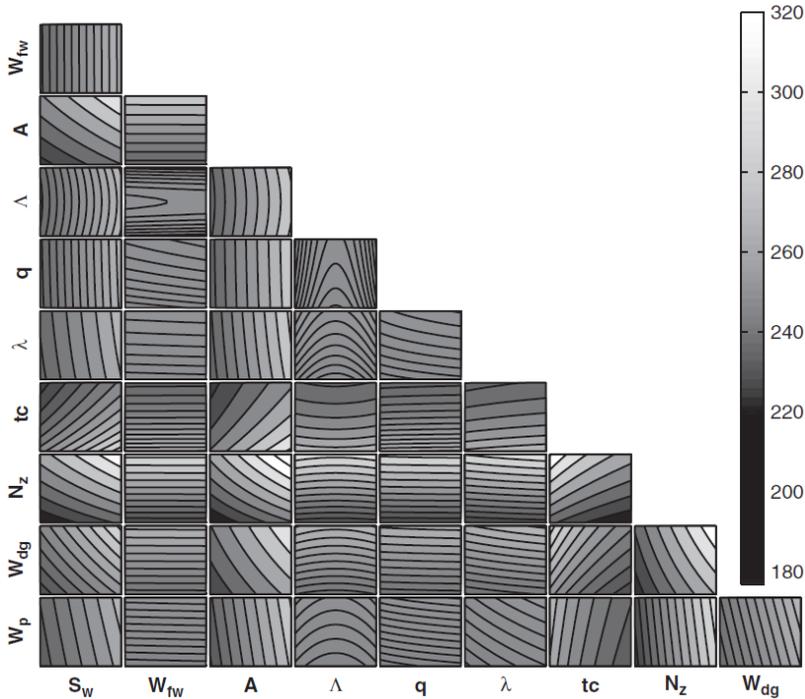


Figura 1. Exemplo de gráficos de contorno em um estudo de Screening.

Fonte: Forrester, Sóbester e Kean (2008)

Onde:

- Sw é a área da asa;
- Wfw é o peso do combustível na asa;
- A é a razão de forma;
- Λ é o quarto de corda sweep;
- q é a pressão dinâmica de cruzeiro;
- λ é a conicidade;
- tc é a espessura do aerofólio em razão da corda;
- Nz é o fator de carga;
- Wdg é o peso bruto da aeronave;
- Wp é o peso da tinta por metro quadrado.

Da análise deste *screening*, os autores destacaram que o peso de tinta por metro quadrado (W_p) não tem influência no formato da asa, onde o fator de carga (N_z), que determina a magnitude da carga aerodinâmica, é muito mais influenciado pelos outros parâmetros. Na Figura 2 são apresentadas as médias e desvios padrão estimados da distribuição dos efeitos de cada uma das dez variáveis do exemplo da asa de avião.

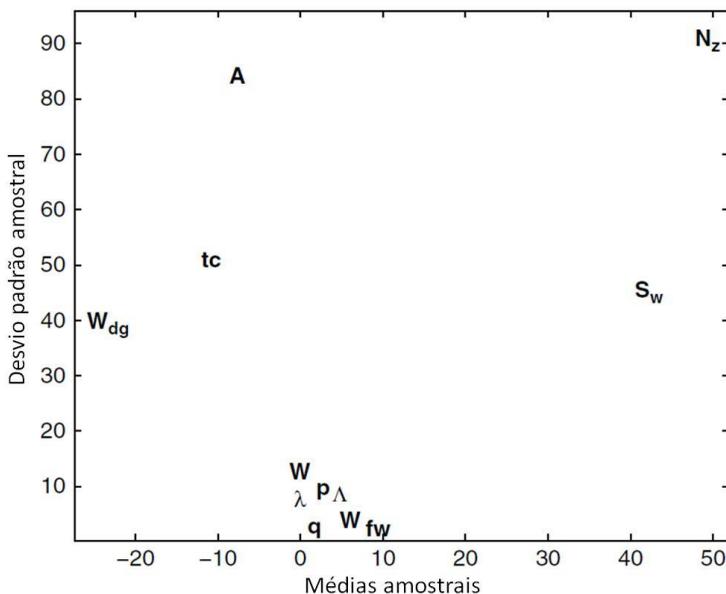


Figura 2. Média e desvios padrão do efeito da distribuição das variáveis.

Fonte: Adaptado de Forrester, Sóbester e Kean (2008)

A primeira observação é que existe um grupo claramente definido de variáveis agrupadas em torno da origem, sendo que a menor medida de tendência central indica as entradas com pouco impacto sobre a função. O peso da pintura (W_p), como esperado, encontra-se entre as variáveis menos influentes. Embora ainda perto da média zero, a variável com maior tendência central dentro deste grupo é o peso de combustível

(Wfw). Sua amostra tem um desvio padrão muito baixo e uma média um pouco maior do que o resto do grupo, o que indica que é mais importante do que elas, mas não está envolvida em interações. O contexto indica que a espessura do aerofólio (t_c) tem importância similar, mas com um efeito não linear. A área da asa (S_w) e o fator de carga (N_z) são as variáveis com impacto mais significativo sobre o peso da asa, porém, os desvios padrão elevados demonstram sua interdependência com os outros parâmetros. O estudo busca demonstrar que com tal conhecimento prévio é possível identificar as variáveis mais importantes para o estudo.

2.2.2. AMOSTRAGEM

Técnicas de amostragem são procedimentos estatísticos para melhorar a representatividade de uma amostra, especialmente quando a obtenção de resultados requer altos investimentos em tempo e dinheiro. O objetivo da utilização de técnicas de estudo amostral é reduzir a variância da estimativa da média dos casos da amostra.

O uso de técnicas de amostragem em simulações de energia em edificações é relevante quando a simulação energética de um universo de casos pode ser muito trabalhosa e qualquer redução no número de simulações para uma análise representa uma diminuição do esforço computacional necessário (MACDONALD, 2009)

Amostragem aleatória simples é uma técnica de amostragem básica. O método funciona através de sorteios ou com a criação de números aleatórios e escalonados para a variável de interesse através da sua distribuição de probabilidade. Todos os elementos da população têm a mesma probabilidade de pertencerem à amostra.

A amostragem estratificada representa uma melhoria em relação à amostragem aleatória simples. Nesta técnica, a distribuição das amostras é dividida em estratos, onde se identificam valores comuns entre os parâmetros do estudo. Em cada estrato é realizada a amostragem aleatória simples, mantendo a proporção das amostras de cada estrato para a dimensão da amostra desejada.

A amostragem por Hipercubo Latino, por sua vez, é uma evolução da amostragem estratificada de Monte-Carlo. O método seleciona valores aleatoriamente, porém dentro de uma distribuição de faixas previamente distribuídas, selecionando um valor aleatório pertencente a cada uma das faixas. Van Dam et al. (2007) explicam o método do hipercubo latino a partir da seguinte ideia: considerando o tabuleiro de xadrez como uma matriz quadrada de 8×8 (quadrado latino¹ de Euler) é possível colocar 8 torres de tal modo que nenhuma torre ataque qualquer outra. Nesse caso, existe uma torre em cada coluna e cada linha. As posições das oito torres constituem uma amostragem por quadrado latino do tabuleiro de xadrez, no sentido que $n = 8$ pontos têm a propriedade de cada linha e cada coluna amostradas por um dos pontos. Generalizando, permite-se ao espaço amostral ter uma dimensão arbitrária M . É possível solicitar que os dados de amostragem caiam exatamente sobre os pontos médios de N intervalos iguais em que se divide o intervalo de coordenadas em cada dimensão. Pode-se, também, simplesmente solicitar que um ponto de amostragem caia em algum lugar entre os intervalos iguais de N . Ainda, pode-se dividir o intervalo em $N-1$ subintervalos iguais, e solicitar que os dados

¹ Um quadrado latino de ordem n é uma matriz $n \times n$ preenchida com n diferentes símbolos de tal maneira que ocorrem no máximo uma vez em cada linha ou coluna.

caiam exatamente nas extremidades destes subintervalos. Neste caso, garante-se que os valores extremos do intervalo serão amostrados. O Hipercubo Latino baseia-se em uma generalização do quadrado latino para um número qualquer de dimensões, por isso do nome hipercubo.

Macdonald (2009) analisou três técnicas de amostragem para análises de sensibilidade para o método Monte-Carlo: amostragem aleatória simples, estratificada e por hipercubo latino. Concluiu que os métodos mais elaborados de amostragem estratificada ou por hipercubo latino não servem para reduzir o número de simulações, porém, são mais robustos quanto à variância da média dos resultados. Para análises de incertezas não há necessidade de utilizar os métodos estratificados ou por hipercubo latino. Entretanto, para um mesmo número de simulações, o Hipercubo latino é mais robusto que os outros métodos.

2.2.3. MÉTODOS DE INFERÊNCIA ESTATÍSTICA

Os modelos estatísticos têm por objetivo fazer afirmações a partir de uma amostra de valores representativos de um universo. Existem várias técnicas estatísticas utilizadas em modelos. A seguir serão apresentadas as técnicas mais utilizadas em modelos de desempenho energético de edificações, como regressão linear múltipla, redes neurais artificiais, máquina de vetores de suporte e processo gaussiano, bem como novos métodos como o de regressão adaptativa multivariada por splines e Randon Forest.

2.2.3.1. REGRESSÃO LINEAR MÚLTIPLA

Regressão linear múltipla é uma técnica estatística multivariada para explorar correlações lineares entre duas ou mais variáveis preditoras para uma única variável dependente. É uma técnica simples e rápida para modelar a predição de desempenho energético de edificações para cargas de refrigeração e aquecimento (KOROLIJA et al., 2013b). A estrutura de uma regressão linear múltipla mais comum é basicamente representada por $y = b_0 + b_1.x_1 + b_2.x_2 + b_3.x_3 + \dots + b_n.x_n$, onde a variável dependente, y , é função de várias variáveis preditoras, x . Os coeficientes b_1, b_2, \dots, b_n , são as constantes da equação, e representam o declive ou acréscimo na variável dependente provocado pelo acréscimo de cada uma das variáveis preditoras, x_1, x_2, \dots, x_n (OGLIARI, 2011). Ao se elaborar um modelo de regressão linear múltipla, pode-se, estatisticamente, incluir todas as variáveis independentes desejadas, ou incluir apenas variáveis sistematicamente selecionadas. Elas podem ser selecionadas em função da significância do parâmetro de regressão de cada uma das variáveis, ou quando se obtém uma melhoria no coeficiente de determinação, R^2 , com sua inclusão.

Modelos de regressão linear múltipla foram adotados no desenvolvimento do RTQ-C e RTQ-R.

2.2.3.2. REDES NEURAIIS ARTIFICIAIS

As Redes Neurais Artificiais (RNA) são sistemas computacionais baseados no funcionamento dos neurônios do cérebro humano. Elas simulam o comportamento de seus receptores de entrada e conexões com outros neurônios para produzir uma saída (HAYKIN, 2001). Redes

Neurais Artificiais são compostas por elementos simples ligados em paralelo. O resultado da RNA depende de sua estrutura, chamada de arquitetura, que está relacionada ao número de camadas de neurônios na camada interna e interações do comportamento entre a entrada e saída de dados através de pesos sinápticos. O número de camadas de neurônios na camada interna deve ser otimizado dependendo do tipo de estudo realizado. Um neurônio, ou nó, é o elemento de processamento básico de uma rede neural e executa funções de soma para determinar a ativação e o resultado de saída do neurônio. O treinamento, ou aprendizado da rede, é armazenado como um conjunto de pesos de conexão entre os neurônios. A rede é treinada diversas vezes para reduzir seu erro, um conjunto de padrões de entrada para treinamento é um grupo de entrada combinados a padrões de saída conhecidos. Um grupo de dados é lido para produzir uma saída, a qual é comparada com o padrão de saída original, isto é, a saída correta ou desejada. Quando existe diferença, os pesos de conexão são alterados para redução dos erros. Isso é feito repetidamente para diferentes padrões de entrada e saída até que todos os erros estejam dentro da tolerância requerida. A rede treinada poderá então ser validada e utilizada para representar a natureza do estudo (KUMAR; AGGARWAL; SHARMA, 2013)

Várias arquiteturas de RNA vêm sendo usadas para predição de desempenho energético de edificações. A arquitetura da rede é o padrão de conexão entre os neurônios e as camadas da rede, e influencia diretamente sua precisão, pois deve ser adequada à natureza de comportamento do estudo. A arquitetura mais utilizada é a de preceptor multicamada com pré-alimentação e retro propagação. Nesta arquitetura, a saída de alguns neurônios alimentam neurônios da mesma camada e de

camadas anteriores. O sinal percorre a rede em duas direções, com capacidade de representar estados em sistemas dinâmicos. Este tipo de arquitetura foi utilizado por Ekici e Aksoy (2009), que conseguiram prever o consumo de energia de três tipos de edificações conforme diferentes orientações, isolamento e percentual de áreas de janela através de RNA com precisão na predição de 94,8 a 98,5% em relação a resultados simulados.

Além do uso de RNA para predição do desempenho da edificação em relação a seu consumo de energia, o comportamento energético de edificações também é analisado para predição de demanda de sistemas de condicionamento de ar em tempo real. Yang, Rivard e Zmeureanu (2005) desenvolveram uma RNA dinâmica que se adapta a padrões e mudanças inesperadas, e que pode ser usada na predição do consumo de energia em tempo real. Eles compararam dois tipos diferentes de arquitetura de RNA e testaram sua precisão em relação ao coeficiente de variação dos resultados.

Melo (2012) propôs modelos de redes neurais artificiais para substituição dos métodos de predição de regressões lineares múltiplas do RTQ-C alcançando melhorias significativas.

2.2.3.3. MÁQUINAS DE VETORES DE SUPORTE

Máquinas de Vetores de Suporte (MVS) é uma categoria de técnicas de aprendizado comparáveis a redes neurais artificiais sem realimentação que pode ser usada para classificação de padrões e regressão linear e não linear.

A inferência por aprendizagem de máquina adota uma técnica de indução, onde conclusões são tomadas a partir de uma amostra. Dois tipos

de aprendizado indutivo podem ser considerados: supervisionado e não supervisionado. No aprendizado supervisionado deve haver um professor externo que apresente o conhecimento da natureza do estudo por exemplos de entrada e saída conhecidos. O conhecimento da máquina é obtido a partir dos exemplos, tornando-a capaz de produzir saídas corretas para entradas não apresentadas anteriormente. No aprendizado não supervisionado não há um professor, não existem exemplos rotulados. A Máquina aprende a representar ou agrupar as entradas submetidas segundo uma medida de qualidade (LORENA; CARVALHO, 2007)

A ideia principal de modelos de Máquina de Vetores de Suporte é construir um hiperplano como resposta. Os vetores de suporte consistem em um pequeno subconjunto dos dados de treinamento extraído. Dependendo de como este núcleo de produto interno é gerado, pode-se construir diferentes máquinas de aprendizagem, caracterizadas por superfícies não-lineares. Enquanto que o algoritmo de retropropagação é planejando especificamente para treinar um *perceptron* de múltiplas camadas, o algoritmo de aprendizagem por vetor de suporte é de natureza genérica por ter uma aplicabilidade mais ampla (HAYKIN, 2001).

As MVS são adequadas para grande dimensão de dados, sobre os quais outras técnicas de aprendizado podem obter resultados super ou subajustados. Ela pode fornecer um bom desempenho na generalização em problemas de classificação de padrões mesmo não incorporando o conhecimento do domínio do problema.

2.2.3.4.PROCESSO GAUSSIANO

O Processo Gaussiano, ou Krigagem², é um conjunto de rotinas de regressão linear que minimizam a estimativa de variância de um modelo de covariância predefinido. Baseia-se no pressuposto de que um parâmetro a ser interpolado pode ser tratado como uma variável regionalizada em um plano de distribuição normal. Uma variável regionalizada é intermediária a uma variável aleatória e uma variável determinista que varia de um modo contínuo a partir de um local para o outro e, por conseguinte, os pontos que estão próximos uns dos outros tem um certo grau de correlação espacial, mas os pontos que são amplamente separados são estatisticamente independentes (MACKAY, 1998). Simplificando, no processo Gaussiano, resultados podem ser presumidos através de uma distribuição normal (distribuição Gaussiana) a partir de pontos conhecidos em um espaço.

2.2.3.5.REGRESSÃO ADAPTATIVA MULTIVARIADA POR SPLINES

Regressão adaptativa multivariada por splines, ou em seu nome original em inglês *Multivariate Adaptive Regression Splines (MARS)*, é uma técnica de regressão multivariada de função não linear apresentada por Friedman (1991). MARS é uma técnica relativamente nova em comparação a outras, apresentando exatidão e redução do custo da computação como principais vantagens (PINA, 2010). Pode ser utilizada tanto para classificação quanto para regressão. É muito útil para

² Nome em homenagem à seu desenvolvedor, D.G.Krige, engenheiro de minas Sul-Africano.

problemas de elevada complexidade e pode ser entendida como uma extensão da regressão linear múltipla, entretanto utiliza estratégias de determinação aleatória de pesos como valores substitutos na elaboração de modelos, estratégia semelhante à utilizada na elaboração de redes neurais artificiais.

O método busca aproximar relacionamentos complexos por regressões lineares em série para diferentes intervalos dos limites das variáveis predictoras ou de subregiões de seus valores. Sua flexibilidade está relacionada a seu poder de adaptação a qualquer forma funcional. Para predições, as RAMS usam funções two-sided truncada, ou de abordagem iterativa forward-backward. Um modelo de MARS pode ser postulado conforme mostram as Equações 2 e 3.

$$\hat{y} = \sum_{n=1}^N a_n \cdot B_{kn}(x_{v(k,n)}) \quad (2)$$

Sendo:

y é a variável dependente;

a_n é o coeficiente de expansão;

B_{kn} é a função da base, definida como:

$$B_n(x_{v(k,n)}) = \prod_{k=1}^k h_{kn} \quad (3)$$

Onde:

x é a variável predictoras;

k é a ordem da interação;

n é o número de termos.

Para $k = 1$ (primeira ordem de interação), o modelo é aditivo, e para $k > 1$, o modelo é interativo par-a-par (FRIEDMAN, 1991). Os modelos MARS buscam representar todo o espaço dos resultados, e durante esta

definição são adicionados um crescente número de funções de base, quando o algoritmo de MARS determina automaticamente as variáveis preditoras mais importantes, assim como a mais significativa interação entre as variáveis preditoras e variáveis dependentes (KINGRE, 2004).

2.2.3.6. RANDOM FOREST

Random Forest (RF), ou Floresta Randômica, é um método de regressão e classificação relativamente novo, apresentado por Breiman (2001). Este método de modelagem combina o resultado de vários modelos de árvores de decisão. Neste método, cada árvore de decisão é construída com profundidade máxima, isto é, sem poda, dentro dos jargões da técnica. Árvores de decisão são métodos de aprendizagem e tomada de decisões baseadas em cadeias de observações de uma amostra. O método de Random Forest permite uma grande diversidade de variáveis para suas interações, isto é, o conjunto-domínio das variáveis não está restrito ao conjunto dos números reais R , como no método SVM, onde o conjunto-domínio de uma variável pode ser um conjunto de dados categóricos ou um conjunto numérico qualquer.

2.3. APLICAÇÃO DE MODELOS DE EDIFICAÇÕES

Modelos de predição de desempenho energético de edificações são utilizados para diversos fins. Estes modelos variam sua complexidade de acordo com a aplicação a qual foram desenvolvidos. Modelos simplificados são utilizados para predição de cargas térmicas para dimensionamento de sistemas, para políticas de eficiência energética de

edificações em certificações, códigos e regulamentos e para ferramentas no desenvolvimento de projetos arquitetônicos e de engenharia.

2.3.1. MODELOS PARA CÁLCULO DE CARGA TÉRMICA

Mao, Haberl e Baltazar (2013b) traçaram o histórico dos modelos de energia desde as definições das leis de gás, estudos de transferência de calor e da termodinâmica. Esta última é responsável pela consolidação dos conceitos de calor, trabalho e energia envolvidos nos cálculos e definições de equações de sistemas térmicos. Estes avanços combinados com as leis de gases e transferência de calor se tornaram a fundação dos cálculos de pico de carga térmica de aquecimento e refrigeração e cálculos de uso anual de energia. Estes cálculos são utilizados para o dimensionamento de sistemas de HVAC. Os autores então descrevem o histórico do desenvolvimento dos cálculos de carga térmica nos E.U.A. desde 1945 até os tempos atuais.

Segundo Mao, Haberl e Baltazar (2013b), o primeiro modelo de cálculo de carga térmica para edificações foi o *Response Factor Method* (RFM), concebido em 1967, e que posteriormente integrou o método da ASHRAE *Handbook of Fundamentals* de 1981, chamado *Weighting Factor Method* (WFM). O método da ASHRAE estimava a demanda de energia através de parâmetros como ganhos térmicos e temperatura da zona térmica. O cálculo do balanço de calor através das leis de conservação de energia para cada ponto que representava uma superfície da zona térmica também foi lançado no mesmo volume de 1981 da ASHRAE *Handbook of Fundamentals*.

Outros métodos de cálculo de carga térmica foram sendo desenvolvidos para serem calculados manualmente, entretanto também surgiram outras formas de predição da carga térmica, como os programas de simulações computacionais e métodos estatísticos orientados por dados (*data-driven*) (MAO; HABERL; BALTAZAR, 2013a), que exploram mais as funcionalidades dos avanços tecnológicos. Muitos dos métodos de simulação usam os modelos de cálculo de carga térmica em seus algoritmos.

2.3.2. MODELOS EM CERTIFICAÇÕES, CÓDIGOS E REGULAMENTOS

A investigação de desempenho energético em edifícios teve início na década de 1920, quando engenheiros mecânicos utilizaram algoritmos de modelagem de fluxo de ar para estudos de sistemas de condicionamento ambiental. A prática evoluiu para predição de cargas de pico de resfriamento em 1950. Após a crise do petróleo da década de 1970 a eficiência energética e a conservação de energia tornaram-se uma prioridade e, auxiliada pelos avanços em informática, a previsão do desempenho energético em edifícios ganhou impulso (RMI, 2010).

Em 1977 o governo sueco lançou um projeto para estimar a economia de energia em residências com base nas suas características construtivas (SANTAMOURIS, 2005). Em 1979 foi desenvolvido um programa computacional para averiguação do consumo de energia e eficiência energética em edificações pelo estado de Montana, Estados Unidos (COMPUTERWORLD, 1979 apud LEIPZIGER, 2013). Novos métodos de avaliação de energia continuaram a ser desenvolvidos desde então.

Depois disso, sistemas de avaliação de eficiência e consumo de energia começaram a surgir. O primeiro sistema de classificação de energia foi desenvolvido em 1981, nos Estados Unidos, pela “*National Shelter Industry Energy Advisory Council*”, que se tornou a RESNET³ em 1995. Ao longo de 1980 e 1990 o *Building Research Establishment* (BRE) realizou dezenas de auditorias energéticas em todo o Reino Unido, que serviu para produzir o “*Domestic Energy Model*”, um antecessor da atual classificação energética BREEAM, lançada em 1990. O BREEAM, por sua vez, é considerado a origem de diversas certificações “verdes”, até mesmo do LEED⁴. E em 1995, nos Estados Unidos, a EPA⁵ criou a ferramenta de classificação *EnergyStar* para edifícios.

O conceito de políticas públicas para classificação de eficiência energética foi concebido na Dinamarca e, embora lento no início, o conceito se espalhou por todo o mundo. A Dinamarca lançou seus sistemas de classificação de eficiência energética obrigatória para edifícios comerciais e residenciais em 1992 e 1993, respectivamente (LEIPZIGER, 2013). Em 2002 a União Europeia lançou sua Diretiva de Desempenho Energético de Edificações - IDEAL/EPBD, uma diretiva que estabeleceu aos Estados membros da União Europeia a implantação de um sistema de classificação energética para informar aos cidadãos

³ RESNET (*Residencial Energy Services Network*) é um sistema de classificação de eficiência energética dos Estados Unidos para edificações residenciais.

⁴ LEED (*Leadership in Energy and Environmental Design*) é um sistema internacional de certificação ambiental para projetos, obras e operações de edificações mais sustentáveis.

⁵ Agência de Proteção Ambiental dos Estados Unidos, criada com a finalidade de proteger a saúde humana e o meio ambiente.

sobre a qualidade termo-energética dos edifícios para construção, venda ou aluguel dos mesmos. Esta diretiva provocou o surgimento de políticas de classificação energética em trinta e um países europeus (BACKHAUS; TIGCHELAAR; BEST-WALDHOBER, 2011). Na década atual, Brasil, China, Austrália, Estados Unidos e outros 24 estados e cidades também aprovaram leis e regulamentos de classificação de eficiência energética para edificações.

Os órgãos reguladores da geração, distribuição e consumo de energia (governo, agências de energia, autoridades locais, etc.) utilizam-se de regulamentos, códigos e certificações para incentivar a economia de energia e maximizar a eficiência energética de edificações. Os regulamentos ou códigos de energia normalmente estabelecem os requisitos mínimos de desempenho para elementos da envoltória de edifícios novos e, em alguns casos, também são estabelecidos requisitos mínimos de eficiência para equipamentos de condicionamento de ar. Já as certificações energéticas de edificações abrangem os procedimentos que permitem a determinação da qualidade de um edifício em comparação a outros, considerando o consumo de energia (PÉREZ-LOMBARD et al., 2009). Para tal, as certificações adotam modelos de predição de desempenho em seus métodos. Estes modelos variam de modelos simplificados à complexos modelos de simulação computacional.

Para certificações energéticas geralmente são utilizados métodos de *Benchmarking*, *Ratings* ou Etiquetação. O primeiro consiste, basicamente, da comparação de um índice de desempenho energético de um edifício com uma amostra de edifícios similares, normalmente com o consumo elétrico mensurado. O *Rating* é uma forma de avaliação da classificação da qualidade energética da edificação através de pontuações.

E a etiquetagem energética de edifícios consiste em atribuir uma classificação de desempenho energético de uma edificação em uma escala de desempenho. No caso da Etiquetagem, a definição da escala de desempenho é mais uma questão política do que técnica (PÉREZ-LOMBARD et al., 2009). Para estes três tipos de certificações o uso de modelos de predição são essenciais. Para desenvolvimento de Benchmarkings, modelos de simulação devem ser adotados para validação dos dados medidos (BORGSTEIN; LAMBERTS, 2014). Já para *Ratings* e etiquetagem, modelos são frequentemente adotados como métodos de simulação ou métodos simplificados para predição do desempenho de edificações.

No Brasil, os dois regulamentos publicados pelo INMETRO, com os Requisitos Técnicos da Qualidade para o nível de eficiência energética em edifícios comerciais, de serviço e públicos (RTQ-C) e para edifícios residenciais (RTQ-R) adotam modelos estatísticos do tipo caixa-cinza, com equações de regressão linear múltipla para avaliação da envoltória das edificações.

2.3.3. MODELOS COMO FERRAMENTAS DE PROJETOS

Tomar decisões em projeto tem como premissa principal a habilidade de predizer o desempenho das alternativas consideradas. Em projeto de edificações isso só é possível através de simulações. As simulações no passado eram limitadas a estudos em croquis, desenhos, maquetes físicas e cálculos manuais (PAPAMICHAEL; PORTA; CHAUVET, 1997), ou na mente e imaginação dos projetistas. Entretanto, o desenvolvimento tecnológico vem permitindo o surgimento de uma

grande variedade de modelos computacionais para simular o desempenho de edificações, conquistando importância significativa nas tomadas de decisões no desenvolvimento de projetos. Programas computacionais para avaliação do desempenho de estruturas, iluminação, energia, conforto térmico, consumo e geração de eletricidade são comuns como ferramentas em diferentes etapas de projetos de edificações.

Kalay (1999) discute a importância do entendimento do desempenho das edificações para o processo de projeto e propõe um processo de projeto orientado pelo desempenho. Ele sugere que as forças por trás de todo processo de projeto estão no desejo de alcançar soluções de qualidade para uma combinação particular de forma e função dentro de um contexto específico. Também sugere que esta qualidade é determinada pela avaliação de desempenho de multicritérios que podem ser integrados em ferramentas computacionais. Pedrini (2003) defende o uso de ferramentas computacionais de análise de desempenho energético nas primeiras fases de projeto. Entretanto, assim como Souza e Knight (2007), mostra que a maioria das ferramentas tem sido desenvolvidas desconsiderando o entendimento dos resultados por arquitetos.

O uso de modelos de todos os tipos é adotado em ferramentas de auxílio a projeto. Papamichael, Porta e Chauvet (1997) desenvolveram um programa computacional para auxílio às decisões de projeto chamado BDA (*Building Design Advisor*), que constava de uma interface gráfica permitindo simular diferentes cenários para uma edificação. Depois, em 2005, Reichard e Papamichael (2005) propuseram a integração deste programa a uma interface com modelos de avaliação de códigos de energia, ajudando nas tomadas de decisões.

Para simulações do desempenho de edificações durante as primeiras etapas de projeto, Kim e Degelman (1998) propuseram o programa CABD. Marsh (1999) também desenvolveu diversas ferramentas para auxílio no processo de projeto que posteriormente foram integradas ao programa ECOTEC. Posteriormente, o ECOTECT (AUTODESK, 2011) foi comprado pela empresa Autodesk que vem trabalhando no desenvolvimento do programa VASARI (AUTODESK, 2014).

Análises de desempenho de edificações a partir de modelos de simulação computacional tem grande potencial para apoiar as decisões no processo de projeto, entretanto, trabalhar com estes modelos pode acrescentar grandes custos ao processo. Mahdavi e El-Bellahy (2005) enfatizam que o custo mais importante não está na aquisição de programas, mas no tempo necessário para aprender e aplicar suas funções. Por isso, modelos simplificados que exijam menos entradas de dados e que deem respostas satisfatórias para tomadas de decisões são propostos para superar esta dificuldade.

Kalogirou e Bojic (2000) propõe o desenvolvimento de modelos de redes neurais artificiais que podem ser usadas efetivamente em predições de desempenho na tomada de decisões dos projetistas. Considerando ser economicamente mais viável avaliar o consumo de energia de uma edificação sem ter que construir um experimento e modelos caros. A aplicação de redes neurais demonstraram que é possível modelar um sistema com o mínimo de entrada de dados e prover aos arquitetos um sistema flexível para testar várias opções rapidamente. A redução do tempo de resposta é o maior benefício deste método. O uso de modelos estatísticos para auxílio no processo de projeto também foi

explorado por Kim, Stumpf e Kim (2011), que analisaram a eficiência energética de edificações através de mineração de dados, que permite descobrir padrões úteis em resultados de simulação para melhorar a eficiência energética dos projetos em desenvolvimento.

2.4. INCERTEZAS NAS MODELAGENS

Todo tipo de modelo adotado para avaliação do desempenho de edificações sempre contará com incertezas. Na avaliação de desempenho de edificações pode-se destacar três tipos delas: incertezas de especificação; incertezas de modelagem; e incertezas de condições de análise. A primeira, pertinente à especificação, tem relação com a ausência de informações das propriedades exatas dos elementos da edificação e informações de uso e ocupação. Estas incertezas estão relacionadas à geometria de ambientes, a superfícies e aberturas, valores de propriedades de seus diversos materiais construtivos e das cargas internas de equipamentos, iluminação e ocupação de pessoas. Já as incertezas de modelagem estão relacionadas às simplificações assumidas para representar o comportamento do sistema ao desenvolver um modelo. As incertezas de condições de análise estão relacionadas às condições climáticas às quais a edificação é avaliada, tanto para arquivos climáticos de referência quanto para dias típicos ou dias de projeto.

Análises de incerteza e estudos de sensibilidade em análises de desempenho de edificações fazem parte de várias pesquisas, utilizados para diversas finalidades. São utilizados para apoio em critérios para decisão de projeto (HOPFE; AUGENBROE; HENSEN, 2013) e comparação de desempenho medido e simulado de edificações para fins

de análise e calibração (HEO et al., 2013; MAILE; BAZJANAC; FISCHER, 2012).

Além disso, muitos estudos buscam determinar as incertezas do processo de simulação na avaliação do desempenho termo energético em edificações para melhorar os processos de predição de modelos de simulação (SILVA; GHISI, 2014). Modelos de simulação refletem uma natureza probabilística do mundo real, por isso existe uma diferença significativa entre o desempenho predito por modelos e o desempenho real de um edifício em uso (DE WILDE, 2014). As ferramentas de simulação são modelos físicos complexos que normalmente incluem mais de 10.000 variáveis, e exigem a especificação de valores com segurança para garantir a qualidade da análise (OLSEN; IVERSEN, 2011). O maior desafio para ferramentas de simulação é lidar com a grande variedade de parâmetros e sua complexidade de não linearidade, descontinuidade e incertezas (HOPFE; HENSEN, 2011).

Para vencer estas limitações, os metamodelos apresentam vantagens, compensando a imprecisão intrínseca dos métodos estatísticos com a facilidade de uso e velocidade nos resultados (MALKAWI; WAEGEL, 2013), além de evitar as incertezas na especificação de milhares de parâmetros, restringindo-se a poucos valores de entrada.

2.5. AGILIDADE EM SIMULAÇÕES COMPUTACIONAIS

O progresso tecnológico e as inovações no campo da informática vêm permitindo que as análises de desempenho de edificações também se desenvolvam. Para o processo de simulações computacionais, se destaca o potencial de simulações cada vez mais rápidas e a possibilidade de

simulações paramétricas de inúmeros casos. A funcionalidade de parametrizar simulações pode ser executada tanto em computadores locais, quanto em máquinas virtuais em servidores na nuvem⁶. Esta possibilidade exige outras funcionalidades para se trabalhar com megadados⁷ e na visualização de seus resultados (NEW et al., 2012).

A modelagem, simulação e pós-processamento de um grande número de casos pode ser um processo demorado ou inviável de ser realizado sem um sistema de automação. Para simulações termo energéticas com o *EnergyPlus*, existem algumas interfaces que facilitam o trabalho com simulações em larga escala. A seguir, são apresentadas algumas ferramentas ou interfaces que automatizam esse processo total ou parcialmente para processamento local ou em servidor web, como o objeto *parametrics* do *EnergyPlus*, o programa *JEPlus*, o *OpenStudio* e o *EnergyPlus Cloud*.

Como ferramenta própria, o *EnergyPlus* possui uma funcionalidade de pré-processamento para combinações paramétricas de variáveis. Esta função permite a criação de uma série de arquivos de entrada para simulação (IDFs), resultantes de um único arquivo de entrada que contenha os objetos e expressões para criação de simulações paramétricas. O pré-processador de paramétricas lê a fonte de entrada e remove os objetos da classe *Parametrics*. Processa estes objetos nas

⁶ O conceito de nuvem em computação refere-se à utilização da capacidade de cálculo, memória e armazenamento de computadores e servidores compartilhados e interligados por meio da Internet.

⁷ Em tecnologia da informação o conceito de megadados (“big data” em inglês) refere-se à análise de um grande banco de dados para geração de resultados interpretáveis.

expressões inseridas no modelo e produz uma série de IDFs para simulação (ENERGYPLUS, 2013b). A função *parametrics* do *EnergyPlus* permite alternar valores específicos de qualquer objeto do modelo. Por exemplo, é possível definir que a espessura de um isolante seja uma variável definida pelo controle do *parametrics*. Além disso, é possível definir expressões aritméticas e trigonométricas para os valores de entrada nas variáveis. Por exemplo, as coordenadas de uma superfície podem ser valores que são definidos de acordo com uma operação aritmética em função de valores definidos. Assim como, a largura de um elemento de sombreamento pode ser definida através da função trigonométrica de um ângulo definido como variável em *parametrics*.

O JEPlus é uma interface desenvolvida para programação e gerenciamento de simulações paramétricas com o *EnergyPlus*, assim como para agrupar resultados em um único arquivo. O programa também suporta simulações com o TRNSYS. Os projetos de simulações paramétricas são organizados em uma estrutura de árvore que reflete a dependência entre os parâmetros para valores específicos de cada alternativa. A entrada de cada valor pode ser distribuída em funções de ocorrência uniformes, gaussianas, triangulares e discretas. Também possui funções para realização de simulações de amostras por Hipercubo Latino. Além disso, possui funcionalidades de simulações em vários processadores de um computador ou em vários computadores em rede (JEPLUS, 2014). O desenvolvimento desta ferramenta é apresentado por Zhang (2009). Paralelamente, existe um serviço online de simulações paramétricas nas “nuvens” com o JEPlus, desenvolvido por Naboni et al. (2013).

Com o uso do JEPlus, Korolija et al. (2013a) desenvolveram arquétipos de escritórios para simulações de desempenho de energia na modelagem de estoques de edificações e estudos paramétricos. O uso da ferramenta possibilitou uma rápida modelagem das simulações para obtenção dos resultados. Já, Zhang e Korolija (2010) demonstraram que o JEPlus pode facilitar a modelagem e simulação paramétrica de estudos complexos. Eles apresentaram um exemplo com a modelagem e simulação de diferentes sistemas de condicionamento de ar, isolamento, tipos de vidro, controle de iluminação e climas, resultando em mais de trinta mil casos. No exemplo, a automação da ferramenta permitiu realizar tais simulações de maneira conveniente, facilitando a entrada e o pós-processamento dos resultados.

Outra iniciativa de integração de simulações paramétricas ao EnergyPlus está na interface em ambiente OpenStudio (NREL, 2013) através do DAKOTA (*Design Analysis Kit for Optimization and Terascale Applications*), um conjunto de algoritmos em código aberto que permite a interface entre códigos de simulação e um sistema interativo de análises estatísticas (ELDRED et al., 2013). Nestes algoritmos é utilizado o pacote de ferramentas DDACE (*Distributed Design and Analysis of Computer Experiments*) para análises de sensibilidade de superfícies de resposta, para determinar quais os parâmetros mais contribuem na variância de um estudo e para avaliar como os parâmetros interagem uns com os outros.

Além disso, o NREL (2014) vem desenvolvendo a ferramenta de análise paramétrica para o OpenStudio, que permite aos usuários lançarem grandes volumes de simulações para serem processadas na nuvem, através do Amazon Web Server (LONG et al., 2014). Como

iniciativa privada, a Autodesk segue na mesma linha com projeto para rodar o EnergyPlus na nuvem com o EnergyPlus Cloud (AUTODESK, 2015). Uma ferramenta web que permite a simulação com EnergyPlus até 20 vezes mais rápidas do que em um computador desktop comum (GOWRI, 2015).

3. CONSTRUÇÃO DA BASE DE DADOS

Para desenvolver um Metamodelo é necessária uma base de dados que represente o universo de casos a serem avaliados. Carlo (2008) e Melo (2012) adotaram tipos de edificações representativas, variando seus parâmetros construtivos para montar tal base. Entretanto, para atender a análise de avaliação por zonas térmicas, proposta por esta tese, a base de dados para esta predição deve representar as possibilidades construtivas e de uso de uma zona térmica de edificação comercial.

Foi definido um universo de casos que representam as possibilidades construtivas em zonas térmicas de edificações comerciais. A partir deste universo foi realizado um estudo amostral por hipercubo latino para reduzir os casos a serem simulados. Para viabilizar o grande volume de simulações, estas foram realizadas através de um Cluster para simulações com o programa EnergyPlus em processos paralelos.

3.1. MÉTODO: PARÂMETROS SIMULADOS

Para representação do comportamento da zona térmica, foram considerados parâmetros físicos, geométricos e carga térmica interna. Para cada condição ou característica, considerou-se uma abrangência de valores que garantisse os intervalos aceitáveis fisicamente ou possíveis de serem encontrados em edificações comerciais. Não foram considerados valores típicos de edificações construídas no Brasil ou valores de referências da bibliografia, pois o objetivo é gerar intervalos uniformes para que os metamodelos possam reproduzir qualquer combinação de aspectos construtivos. Os valores foram distribuídos homogeneamente para alimentar a produção dos modelos estatísticos permitindo a definição

das funções a partir de pontos equidistantes. Uma distribuição discreta dos valores foi adotada para controlar o intervalo de cada parâmetro.

3.1.1. PARÂMETROS TÉRMICOS

Os parâmetros térmicos são os referentes às propriedades térmicas dos materiais opacos e transparentes considerados para elaboração dos metamodelos. Foram considerados parâmetros de paredes, coberturas e vidros que estão estabelecidos pelas normas e outros regulamentos adotados no país.

3.1.1.1. TRANSMITÂNCIA E CAPACIDADE TÉRMICA

A estratégia de modelagem das paredes foi desenvolvida para permitir a parametrização deste componente para as simulações. Para isso, foram elaboradas paredes compostas de dois materiais. Estas paredes representam diferentes tipologias construtivas apenas variando as espessuras destes materiais. Os materiais foram combinados em duas camadas: concreto e isolante térmico. Foram adotadas três diferentes espessuras para camada de concreto e quatro espessuras para camada de isolante. O resultado de combinações possíveis de espessuras para estes dois materiais seria 12 tipos de parede, porém a combinação de 0,0001m de concreto e 0,00001m de isolante térmico foram descartadas, pois a transmitância térmica e a capacidade térmica não representariam nenhum sistema construtivo real. Assim, foram adotados 11 tipos de paredes com três níveis de transmitância térmica e três níveis de capacidade térmica, como apresentado na Tabela 2

Tabela 2. Tipos de paredes simuladas

Tipo de parede	Concreto (m)	Isolante (m)	Transmitância Térmica (W/m².K)	Capacidade Térmica (kJ/m².K)
PAR1	0,1	0,00001	4,396	220
PAR2	0,1	0,005	2,539	220
PAR3	0,1	0,025	0,943	220
PAR4	0,1	0,05	0,528	220
PAR5	0,04	0,00001	5,176	88
PAR6	0,04	0,005	2,781	88
PAR7	0,04	0,025	0,975	88
PAR8	0,04	0,05	0,538	88
PAR9	0,0001	0,005	2,97	0,22
PAR10	0,0001	0,025	0,997	0,22
PAR11	0,0001	0,05	0,544	0,22

A construção das coberturas foi similar ao processo utilizado para paredes externas. A Tabela 3 apresenta os 11 tipos de coberturas, com 3 níveis de transmitância térmica e três níveis de capacidade térmica.

Tabela 3. Tipos de coberturas simuladas

Tipo de cobertura	Concreto (m)	Isolante (m)	Transmitância Térmica (W/m².K)	Capacidade Térmica (kJ/m².K)
COB1	0,1	0,00001	3,739	220
COB2	0,1	0,005	2,305	220
COB3	0,1	0,025	0,909	220
COB4	0,1	0,05	0,517	220
COB5	0,04	0,00001	4,288	88
COB6	0,04	0,005	2,503	88
COB7	0,04	0,025	0,938	88
COB8	0,04	0,05	0,527	88
COB9	0,0001	0,005	2,654	0,22
COB10	0,0001	0,025	0,958	0,22
COB11	0,0001	0,05	0,533	0,22

As construções das paredes e coberturas foram modeladas com a camada de isolamento voltado para o interior da zona térmica, a fim de reproduzir o efeito de atraso térmico da massa do concreto. A inércia térmica interna é considerada por um parâmetro específico de inércia térmica interna.

3.1.1.2. *ABSORTÂNCIA SOLAR DA PAREDE EXTERNA E COBERTURA.*

Absortância solar representa a fração do calor da radiação solar incidente que é absorvida pelo material. É um valor adimensional que pode variar de 0,0 a 1,0. A variação de valores de absortância solar da parede externa e cobertura foram consideradas individualmente, porém ambos foram variados de 0,2 a 0,8, em intervalos de 0,1.

3.1.1.3. *INÉRCIA TÉRMICA INTERNA*

As superfícies internas da zona térmica foram consideradas adiabáticas e sem massa térmica. A condição adiabática foi usada para que a zona térmica represente uma porção de qualquer ambiente. E foi considerada sem massa térmica para que o efeito da massa térmica interna seja controlado por um objeto independente. Portanto, para variar a massa térmica interna na simulação foi utilizado o objeto *InternalMass* do EnergyPlus, adotando a variação da área de uma parede interna pesada, com 10cm de concreto (220 kJ/m²K). Essa área foi considerada com a variação de 2 a 56m². Estes valores podem representar uma grande variação de massa térmica interna, desde uma massa desprezível de construções leves até um sistema construtivo pesado. Por exemplo, o valor maior, de 56m², equivale a uma zona térmica de 6,22m de altura com as três paredes internas de 10cm de concreto. Neste trabalho foi usada uma variação discreta dos valores de área para massa térmica interna adotando-se um intervalo de dez valores: 2, 8, 14, 20, 26, 32, 38, 44, 50 e 56 m².

3.1.1.4. TIPO DE VIDRO

A simulação detalhada de vidros exige uma complexidade de modelagem que inviabiliza sua parametrização. Para viabilizar o estudo de diferentes combinações de tipos de vidro foi utilizado um método simplificado de simulação de vidros com o objeto *WindowMaterial:SimpleGlazingSystem* do programa EnergyPlus. Para simulação de vidros com este objeto basta especificar os valores de Fator Solar e transmitância térmica do vidro que o programa produz um material com estas propriedades para rodar a simulação.

Para representar vidros claros, com fator solar de 0,87, até vidros de controle solar de alto desempenho, com fator solar de 0,21, foram considerados cinco valores intermediários. Os valores considerados para Fator Solar do vidro foram: 0,87; 0,76; 0,65; 0,54; 0,43; 0,32; e 0,21.

O insulamento do vidro foi considerado variando seus valores de transmitância térmica. Foi considerada a transmitância média de vidros simples ($5,7\text{W}/\text{m}^2\text{K}$), vidros duplos ($2,8\text{W}/\text{m}^2\text{K}$) e vidros triplos ou com camada de baixa emissividade ($1,9\text{W}/\text{m}^2\text{K}$).

3.1.2. PARÂMETROS GEOMÉTRICOS

Para modelar a simulação de uma zona térmica a fim de representar um universo de zonas possíveis foi feita a parametrização da geometria. Considerando que, em um balanço térmico, o importante são as proporções, a parametrização foi feita em função da relação das superfícies construídas com a área de piso da zona térmica.

3.1.2.1. PERCENTUAL DE JANELA NA FACHADA

O percentual de janela na fachada é a razão da área envidraçada e a área total de fachada. Este parâmetro tem grande importância nos ganhos térmicos dos ambientes. A variação deste parâmetro foi simulada através do aumento da janela da fachada externa da zona térmica. A Figura 3 ilustra os percentuais de janela na fachada da zona térmica.

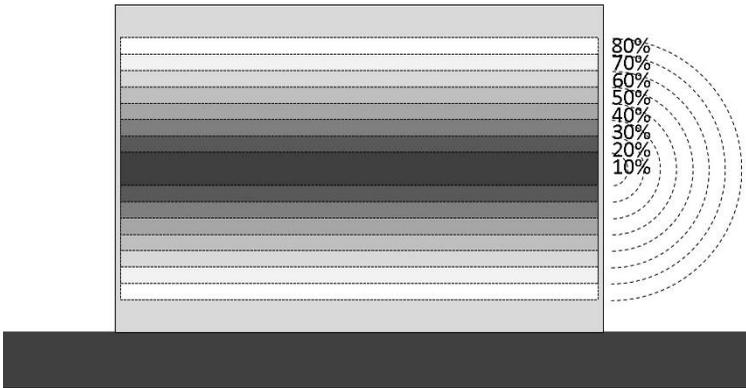


Figura 3. Percentual de janela na fachada da zona térmica

A zona térmica foi simulada desde a situação sem janela até um percentual de janela na fachada de 80%. Sendo adotados os valores de 0, 10, 20, 30, 40, 50, 60, 70 e 80%.

A abertura envidraçada foi parametrizada para variar sua área a partir do centro da parede externa. Porém, é importante ressaltar que este tipo de modelagem só é válido pois as simulações estão avaliando apenas o comportamento termo energético da zona térmica, ignorando a distribuição da radiação e luz solar no ambiente.

3.1.2.2. *SOMBREAMENTO DA ABERTURA*

Existem vários tipos de elementos de sombreamento para aberturas utilizados em edificações comerciais. Sua função principal é bloquear a radiação solar incidente nas áreas envidraçadas e, assim, reduzir os ganhos de calor. Basicamente, a eficiência de um elemento de sombreamento é dada em função de seu mascaramento da trajetória solar sobre as janelas. Neste trabalho foram adotados três ângulos de obstrução de elementos de sombreamento sobre a janela da zona térmica, considerados como os básicos para se definir sua eficiência de obstrução: Ângulo Vertical de Sombreamento (AVS); Ângulo Horizontal de Sombreamento (AHS); e Ângulo de obstrução da vizinhança (AOV). É importante salientar que a orientação solar é outro parâmetro deste estudo, onde a combinação de orientação solar e ângulos de sombreamento representam uma gama de obstruções que pode ser considerada representativa para edificações comerciais. Elementos como venezianas, persianas, telas e placas perfuradas não foram contemplados neste trabalho.

O AVS é o ângulo formado entre o elemento de sombreamento horizontal e a base da janela. Na simulação, ele foi modelado como uma superfície perpendicular a aresta superior da janela projetada para frente, não se estendendo para as laterais além do limite da janela. O AHS são os ângulos formados por elementos de sombreamento na vertical e as arestas direita e esquerda da janela. Ambos os ângulos foram variados entre 0 e 80° e estão representados na Figura 4.

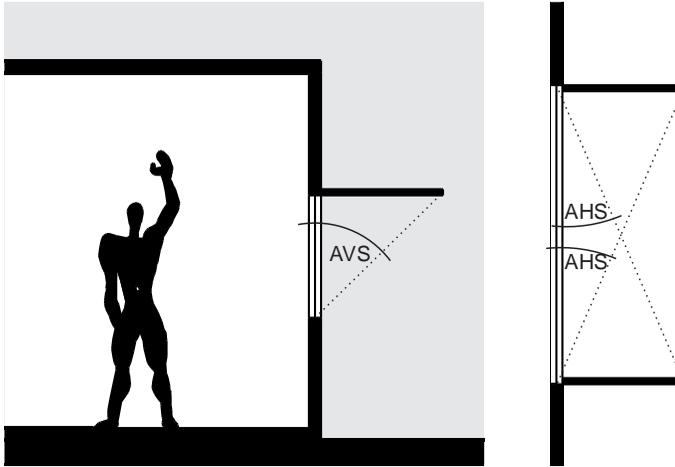


Figura 4. Representação do AVS e AHS em relação a janela.

Para representar o efeito do sombreamento de uma edificação vizinha, foi simulada uma superfície paralela à fachada da zona térmica, variando-se sua altura e largura. Esta superfície foi colocada a cinco metros de distância da fachada, porém o parâmetro de interesse é seu ângulo de obstrução da abóboda celeste. Esta superfície foi simulada sem refletância, para apenas representar o sombreamento vizinho, sem possíveis reflexões. O ângulo que define a proporção altura e largura desta superfície foi chamado de Ângulo de Obstrução Vizinha (AOV). Este ângulo teve seu valor variado entre 0 e 80° e está representado na Figura 5.

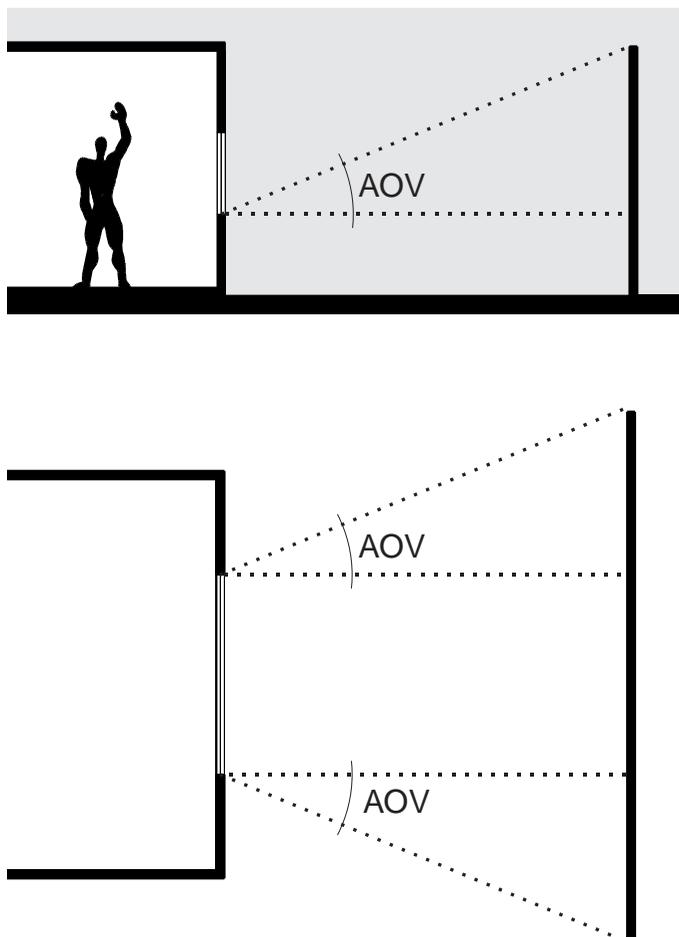


Figura 5. Representação do AOV em relação a janela.

A superfície que representa a obstrução vizinha aumenta sua área exponencialmente conforme se aumenta o AOV. Entretanto, em uma projeção estereográfica, a obstrução desta superfície aumenta de forma a obstruir a abóboda celeste proporcionalmente, conforme ilustrado pela Figura 6.

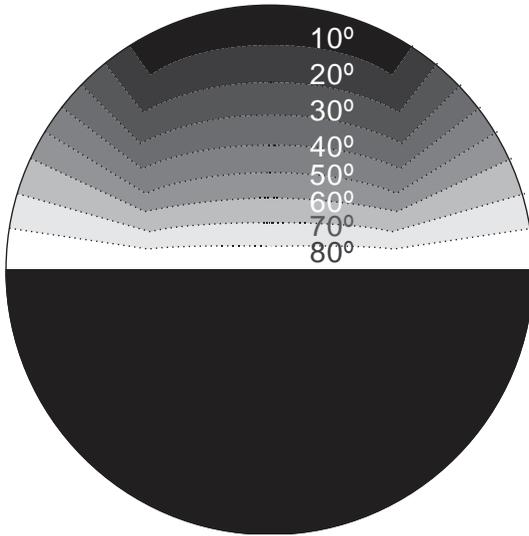


Figura 6. Superfície de obstrução conforme AOV em projeção estereográfica.

Todos os ângulos de sombreamento são representados por elementos fixos e sempre iguais, sem variação para esquerda ou para direita, sendo uma simplificação dos dispositivos de sombreamento que podem formar os respectivos ângulos.

3.1.2.3. EXPOSIÇÃO DO PISO E COBERTURA

Para a zona térmica representar uma zona em uma edificação comercial real, as exposições do piso e teto foram modeladas para representar ambientes em condição de térreo, pavimento intermediário, sobre pilotis e em pavimento de cobertura. Dois parâmetros definem esta situação: Exposição do Piso e Exposição da Cobertura. A Figura 7 representa como as condições de exposição do piso e cobertura definem a situação da zona térmica em uma edificação.

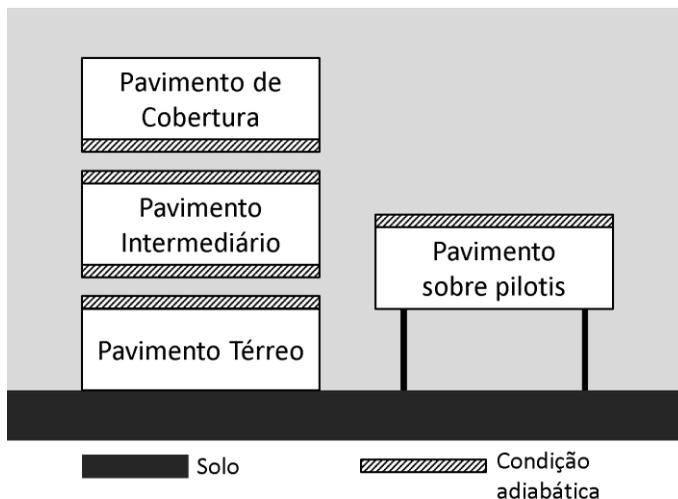


Figura 7. Tipos de exposição do piso e cobertura

Conforme a Figura 7, a exposição do piso pode estar em contato com o solo (representando um pavimento térreo), voltado para o exterior (em situação de pilotis) ou adiabática (quando em pavimento intermediário). A exposição do teto pode estar voltada para o exterior (em pavimento de cobertura) ou em situação adiabática (quando em pavimento intermediário).

3.1.2.4. ORIENTAÇÃO SOLAR

Para que o modelo possa representar a orientação solar de qualquer zona térmica de edificação comercial, a zona térmica simulada foi parametrizada com sua fachada principal voltada para todas as orientações solares em intervalos de 45° . Desta forma, a variável Azimute possui oito valores que representam a orientação da fachada para os quatro pontos cardeais (Norte, Sul, Leste e Oeste) e para os quatro pontos colaterais (Nordeste, Sudeste, Noroeste e Sudoeste).

3.1.2.5. PÉ-DIREITO

Para simular a variação de altura nos ambientes comerciais foram considerados valores de pé-direito de 2,6 a 6,2m. O pé-direito tem muita influência na avaliação de eficiência energética de edificações, pois está diretamente relacionado com o volume de ar dos ambientes e, conseqüentemente, com a energia demandada para condicionar este volume de ar. Os valores de pé-direito adotados foram: 2,6; 3,0; 3,4; 3,8; 4,2; 4,6; 5,0; 5,4; 5,8 e 6,2m. Ao aumentar o pé-direito nas simulações, as variáveis geométricas de área envidraçada e ângulos de sombreamento relacionadas à proporção de altura da zona térmica foram ajustadas automaticamente. Assim como o tamanho da janela, ajustados conforme o percentual de área envidraçada na fachada e os tamanhos dos elementos de sombreamento conforme dimensões das janelas.

3.1.3. CARGA TÉRMICA INTERNA

Os parâmetros de carga térmica interna descrevem os ganhos de calor produzidos dentro da zona térmica. Foram consideradas as cargas térmicas internas do sistema de iluminação, equipamentos e pessoas. A infiltração de ar, apesar de não ser uma carga térmica interna, foi adotada nesta categoria, pois está relacionada a trocas de calor do ambiente interno.

3.1.3.1. DENSIDADE DE CARGA DE ILUMINAÇÃO E EQUIPAMENTOS

As cargas internas do sistema de iluminação e equipamentos foram agrupadas como uma carga única com fração radiante de 72% e fração

convectiva de 10%, o restante da carga representa a fração visível para lâmpadas ou a fração de trabalho de equipamentos. Esta repartição de frações é relativa à carga de uma luminária suspensa com lâmpadas fluorescentes. Este parâmetro foi adotado como densidade de carga em W/m^2 e foi variado de 4 a $40W/m^2$ (sendo 4, 8, 12, 16, 20, 24, 28, 32, 36 e $40W/m^2$). Agrupar as duas cargas permite representar diversas possibilidades de combinações em apenas um valor. Por exemplo, uma carga de $20W/m^2$ pode representar um sistema de iluminação de $15W/m^2$ somado a $5W/m^2$ de equipamentos ou representar um sistema de iluminação eficiente de $9W/m^2$ e $11W/m^2$ de equipamentos.

A carga de iluminação, através da fração radiante, representa um sistema de luminárias com lâmpadas fluorescentes montadas sob o forro e a carga de equipamentos representa equipamentos elétricos como computadores, televisores, e máquinas elétricas que não possuam carga de calor latente. Entretanto, apesar do parâmetro de densidade de carga térmica de equipamentos e iluminação ser simulado como sendo estes tipos de sistemas, a contribuição à carga térmica é que implica o seu uso. Portanto, este parâmetro poderá representar sistemas diferentes.

3.1.3.2. PESSOAS

A densidade de pessoas é considerada como parâmetro para simular o efeito da carga térmica da presença de pessoas no ambiente. O ganho térmico pela ocupação de pessoas é dado por calor latente e sensível. O EnergyPlus calcula automaticamente a divisão destes ganhos. A fração do calor sensível é dividida entre os ganhos de calor radiante e ganhos de calor por convecção. A fração radiante dos ganhos de calor sensível pelas pessoas foi fixado em 30%. Este valor representa a fração

de energia emitida pelas pessoas por radiação de ondas longas. O complemento deste valor é assumido como carga térmica por convecção, ou seja, 70%. O total de ganho de calor por pessoa é definido por seu nível de atividade física, este nível de atividade foi fixado em 120W por pessoa, equivalente a uma atividade média de escritório. Da carga térmica por pessoas, foi variada apenas a densidade de pessoas na zona térmica. Foi adotada uma variação de 0,1 a 0,7 pessoas por metro quadrado. Essa variação pode representar a ocupação de alta densidade, como salas de aula ou centrais de telefonia com 0,7 pessoa/m² ou salas de escritório ou consultórios médicos de baixa ocupação com 0,1 pessoa/m². Este intervalo foi dividido em seis valores: 0,1; 0,2; 0,4; 0,5; 0,6; e 0,7 pessoa/m².

3.1.3.3. *INFILTRAÇÃO*

Infiltração é o fluxo de ar indesejado ou acidental que vem do ambiente externo para o interno. Geralmente a infiltração é causada por frestas em portas e janelas ou pela operação de abertura e fechamento de portas de acesso. Para simular o efeito da infiltração, foi adotado o modelo simplificado, com o qual se especifica uma taxa de infiltração deduzida. A especificação de valores de taxa de infiltração foi em “trocas de ar por hora”, onde é considerada a renovação do volume de ar da zona térmica pela infiltração de ar do ambiente externo. Para a construção da base de dados foram considerados os valores: 0,2; 0,4; 0,6; 0,8; e 1,0 trocas de ar por hora.

3.1.3.4. HORÁRIO DE OCUPAÇÃO

O horário de ocupação e de controle de iluminação e equipamentos influencia os ganhos por cargas térmicas internas. Para este trabalho foram adotados padrões de horários que controlam simultaneamente a ocupação de pessoas, equipamentos, iluminação e infiltração. Como simplificação, as horas de ocupação ou uso são consideradas com 100% da carga de cada ganho térmico interno. Para os dias de semana foram considerados horários de 8 a 24 horas de ocupação (Figura 8) e os fins de semana foram considerados sem ocupação.

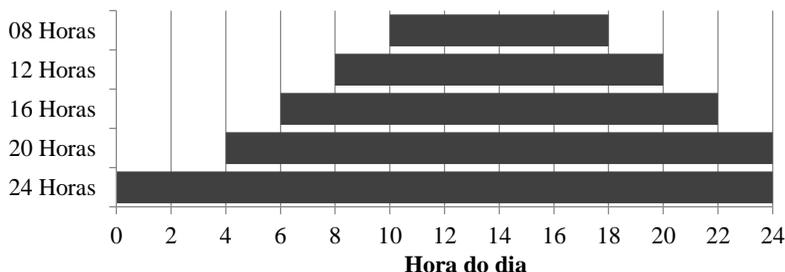


Figura 8. Horários de ocupação dos dias de semana.

Conforme a Figura 8, foram considerados horários a partir de uma ocupação de 8 horas, com a variação de quatro em quatro horas até um horário de ocupação de 24 horas.

3.1.4. TOTAL DE PARÂMETROS

A seleção dos parâmetros físicos, geométricos e de carga térmica interna resultou em um conjunto de 21 variáveis de 3 a 11 valores cada, conforme apresentado na Tabela 4. A Figura 9 representa a zona térmica e seus parâmetros variáveis.

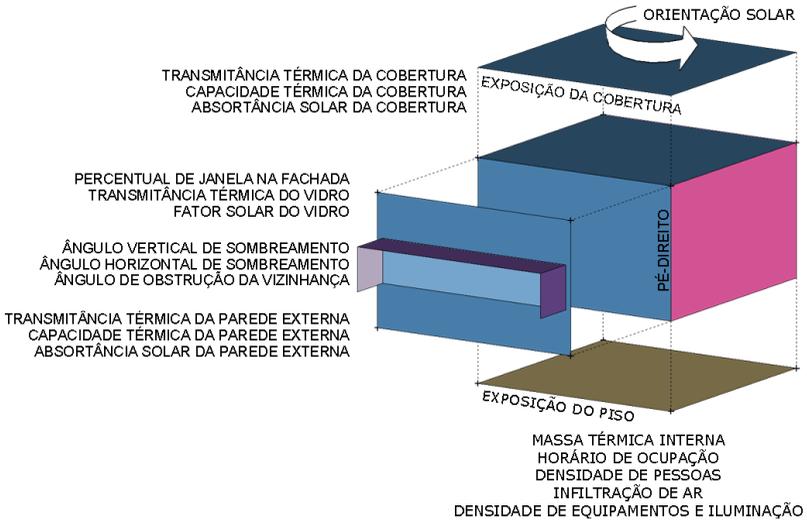


Figura 9. Representação da zona térmica e seus parâmetros variáveis

Tabela 4. Parâmetros e valores simulados

Parâmetro	Un.	Valores
Massa Térmica Interna	m ²	2; 8; 14; 20; 26; 32; 38; 44; 50; 56
Transmitância Térmica da parede externa por tipo de parede (Upar)	W/m ² .K	4,396(PAR1); 2,439(PAR2); 0,943(PAR3); 0,528(PAR4); 5,176(PAR5); 2,781(PAR6); 0,975(PAR7); 0,538(PAR8); 2,997 (PAR9); 0,997(PAR10); 0,544 (PAR11)
Capacidade Térmica da parede externa por Tipo de parede (CTPar)	kJ/m ² .K	0,22 (PAR11, PAR9, PAR10); 88 (PAR8, PAR 7, PAR6); 220 (PAR3, PAR2, PAR1)
Absortância solar da parede externa (aPar)	-	0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8
Transmitância térmica da cobertura por tipo de parede (UCob)	W/m ² .K	3,305(COB1); 2,305(COB2); 0,909(COB3); 0,517(COB4); 4,288(COB5); 2,503(COB6); 0,938(COB7); 0,527(COB8); 2,654 (COB9); 0,958(COB10); 0,533 (COB11)
Capacidade Térmica da cobertura por tipo de parede (CTCob)	kJ/m ² .K	0,22 (COB11, COB9, COB10); 88 (COB8, COB 7, COB6); 220 (COB3, COB2, COB1)
Absortância solar da cobertura (aCob)	-	0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8
Percentual de janela na fachada (WWR)	%	0; 10; 20; 30; 40; 50; 60; 70; 80
Fator Solar do vidro (FS)		0,87; 0,76; 0,65; 0,54; 0,43; 0,32; 0,21
Transmitância térmica do vidro (UVid)	W/m ² .K	5,7; 2,8; 1,9
Ângulo vertical de sombreamento (AVS)	°	0; 10; 20; 30; 40; 50; 60; 70; 80
Ângulo horizontal de sombreamento (AHS)	°	0; 10; 20; 30; 40; 50; 60; 70; 80
Ângulo de obstrução da vizinhança (AOV)	°	0; 10; 20; 30; 40; 50; 60; 70; 80
Horário de ocupação (Ocup)	horas	8h; 12h; 16h; 20h; 24h
Densidade de Equip.e Iluminação (DEI)	W/m ²	4; 8; 12; 16; 20; 24; 28; 32; 36; 40
Densidade de pessoas (DPessoas)	pessoas/m ²	0,1; 0,2; 0,4; 0,5; 0,6; 0,7
Orientação Solar (Azimute)	°	0; 45; 90; 135; 180; 225; 270; 315
Infiltração	ACH	0,2; 0,4; 0,6; 0,8; 1,0
Pé-direito	m	2,6; 3,0; 3,4; 3,8; 4,2; 4,6; 5,0; 5,4; 5,8; 6,2
Exposição da Cobertura (Exp_Cob)	-	Externa; adiabática
Exposição do Piso (Exp_Piso)	-	Contato com o solo; adiabática; externa

A combinação de todos estes parâmetros resulta em mais de cinco quatrilhões de casos (5.881.705.552.800.000), um número inviável de ser simulado e analisado atualmente. Portanto, foi utilizado um método de amostragem para reduzir o número de casos sem afetar significativamente a representação deste universo. A amostragem foi feita por Hipercubo Latino para garantir o equilíbrio da distribuição dos valores de cada parâmetro. Esta amostragem foi processada para um milhão de combinações que, para manter a distribuição dos valores, resultou em 1.293.250 casos.

3.2. MÉTODO: SIMULAÇÕES

Para simular a amostra de aproximadamente 1,29 milhões de casos foi modelada uma zona térmica para simulação com o programa EnergyPlus. Esta zona térmica foi utilizada como base para variação dos parâmetros através de um script elaborado para rodar em um Cluster computacional.

3.2.1. ZONA TÉRMICA

A zona térmica simulada tem 4,5m de largura e 4,5m de profundidade. Sua altura varia conforme o pé-direito do caso. Só há uma fachada voltada para o exterior, as outras são adiabáticas. Na fachada externa há uma janela que varia sua área conforme o WWR do caso. Na janela existem três superfícies perpendiculares que formam os elementos de sombreamento como brises verticais e um brise horizontal, estes elementos variam conforme o AVS e o AHS. Uma superfície a 4,5m de

distância paralela a fachada principal representa o sombreamento de elementos e edificações vizinhas, e sua dimensão varia conforme o AOV. A geometria da zona térmica simulada está representada na Figura 10.

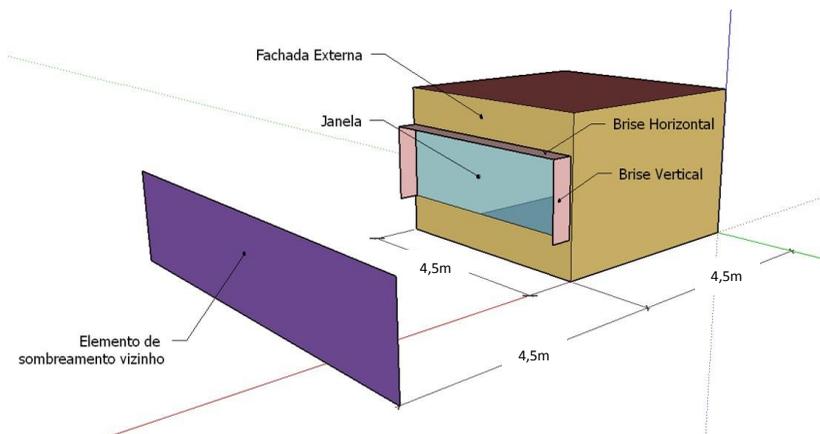


Figura 10. Representação do modelo de simulação

Todos os parâmetros variáveis são manipulados no arquivo de entrada do EnergyPlus (IDF). Esta manipulação é feita por um Script que roda no Cluster de Simulação montado para rodar todas as combinações da amostra.

A fachada com janela é a única fachada voltada para o exterior, as paredes internas são consideradas adiabáticas. Esta condição, apesar de ignorar as trocas térmicas internas da edificação, é uma estratégia comum em simulações computacionais de edificações inteiras, quando a repetição de pavimentos ou de zonas térmicas é representada pela multiplicação numérica de suas relações.

Apesar da perda do efeito de transferência de calor pelas paredes internas, os ganhos de calor da edificação são respeitados. Este mesmo princípio foi adotado para o metamodelo de predição do desempenho de zonas térmicas.

3.2.2. CLUSTER DE SIMULAÇÃO

Um Cluster⁸ foi montado para processar um grande volume de simulações computacionais de estudos paramétricos para análise de desempenho de edificações com o programa EnergyPlus. O desenvolvimento deste Cluster é parte de um trabalho conjunto do Laboratório de Eficiência Energética em Edificações (LabEEE) e o Lawrence Berkeley National Laboratory. Este trabalho teve suporte do Diálogo Estratégico de Energia entre Brasil e Estados Unidos através da colaboração de Geof Sawaya, *Building Technologies Compute Fellow*.

O sistema montado para estas simulações é um conjunto de máquinas virtuais em um servidor rodando Ubuntu (versão 12:04) e com EnergyPlus (versão 8.0) instalado. A paralelização foi feita de duas formas: na distribuição de processos entre várias máquinas virtuais e através de núcleos locais em cada máquina virtual. O sistema possui oito máquinas Intel I7 em um servidor Xeon X5680@3.33GHz com 6 núcleos de 8 GB de memória RAM cada, além de cerca de 1 TB de memória para armazenamento.

Foi desenvolvido um Script para OpenStudio Application, em linguagem Ruby, para manipular os objetos dos arquivos IDF (*Input data file*), arquivo de entrada do EnergyPlus. A maior parte do trabalho foi automatizada, os parâmetros e os valores combinados foram especificados e um IDF da zona térmica (arquivo de entrada de dados do EnergyPlus) foi utilizado como caso base para executar as simulações. O sistema combinou os valores dos parâmetros da combinação paramétrica

⁸ Cluster é um conjunto de computadores agregados para trabalhar como se fosse uma única máquina de grande porte.

sobre o caso base e gerou os arquivos de entrada do EnergyPlus. A gestão de simulação foi programada para proceder de maneira altamente eficiente, mantendo os núcleos de processamento paralelo saturados durante o tempo necessário para completar o conjunto de simulações.

3.2.3. ARQUIVO CLIMÁTICO

Para o estudo, as simulações foram processadas para o arquivo climático TRY (*Test Reference Year*) de Florianópolis (LABEEE, 2014). Este arquivo climático é composto por dados climatológicos registrados e estimados do ano de 1963, do qual foi o mais representativo entre 1961 e 1970 (GOULART; LAMBERTS; FIRMINO, 1998). Este arquivo climático possui limitações nos dados de radiação solar global horizontal e direta normal, pois Carlo e Lamberts (2005) precisaram estimar estes valores para poder utilizá-los em simulações de desempenho energético de edificações. Porém, este arquivo climático TRY de Florianópolis foi adotado por ter os dados mais utilizados e estudados em simulações de desempenho térmico de edificações no Brasil, podendo ser considerado consagrado para estes tipos de estudos.

3.2.4. VARIÁVEIS DEPENDENTES

Como resultados das simulações foram solicitadas as cargas de energia integrada mensais para refrigeração e aquecimento (kWh) e os picos anuais de carga de refrigeração e aquecimento (W). Destes resultados, as cargas de energia integrada mensais para refrigeração foram somadas e normalizadas pela área da zona térmica, resultando na energia integrada anual da carga de refrigeração (kWh/m².ano), doravante

chamada de carga anual de refrigeração, adotada como variável dependente. Este valor foi utilizado como coeficiente de desempenho energético de edificações, pois a energia integrada da carga térmica é diretamente relacionada à eficiência da envoltória sem interferência do sistema de condicionamento de ar. Para computar a energia e os picos de carga térmica, foi simulado um sistema de condicionamento de ar ideal durante as horas de ocupação, que fornece energia para suprir toda carga térmica requerida para manter a zona térmica dentro do intervalo de temperatura estabelecido por um termostato. Este sistema é chamado de ideal porque possui uma capacidade de refrigeração e aquecimento infinita. Nas simulações, o sistema é acionado nos horários de ocupação com termostato de 24°C para refrigeração e 20°C para aquecimento.

3.3. RESULTADOS: APRESENTAÇÃO DA BASE DE DADOS

O trabalho computacional de simulações com EnergyPlus no Cluster levou aproximadamente 150 horas. Como resultado foi gerado um arquivo de base de dados em formato de valores separados por vírgula (csv) com 403 Megabytes. Um arquivo com 1.293.250 de linhas e 69 colunas. Cada linha representando um caso simulado e cada coluna representando uma variável, sendo 21 variáveis que independentes e 48 variáveis de saída. Lembrando que as variáveis de saída da base de dados são os valores mensais de carga e pico de refrigeração e aquecimento.

A frequência de ocorrência de cada valor de parâmetro preditor na amostra é praticamente a mesma devido a amostragem por hipercubo latino. Entretanto, a frequência de ocorrência dos valores das variáveis

dependentes são resultados das simulações. As distribuições dos valores de carga térmica anual e carga de pico de refrigeração e aquecimento são apresentadas na Figura 11.

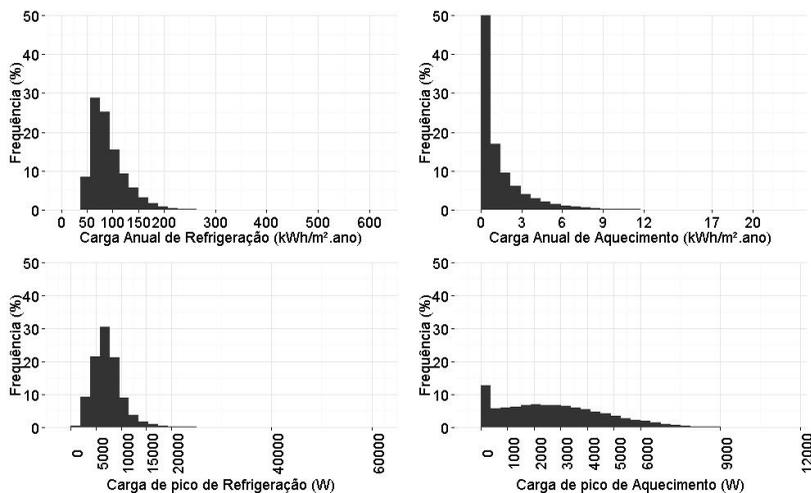


Figura 11. Histogramas de cargas anuais e picos de Refrigeração e Aquecimento

Os valores de carga anual de refrigeração apresentam distribuição enviesada para esquerda com concentração entre 75 e 100kWh/m².ano. O histograma da carga anual de aquecimento resultou em mais de 50% dos casos com menos de 1kWh/m² ano. O pico de refrigeração apresentou comportamento semelhante à carga anual de refrigeração. A carga de pico de aquecimento apresenta uma tendência de distribuição normal, porém com os casos da calda esquerda, que ficariam negativos, concentrados em zero.

Todavia, este trabalho se concentra no estudo da variável dependente de carga térmica anual para refrigeração. A base de dados com os resultados das simulações está apresentada em gráficos do tipo “*violin plot*”, que é a junção do diagrama de caixa (boxplot) com histogramas de

densidade. Os histogramas de densidade são apresentados com as curvas de densidade na vertical dos dois lados da linha central do diagrama de caixa. O diagrama de caixa apresenta a mediana e os quartis inferior e superior. Cada quartil representa 25% dos casos, portanto toda caixa representa metade de todos os casos. As extremidades das linhas que cruzam a caixa representam os valores máximos e mínimos da base de dados, excluindo os valores considerados extremos, que são representados por pontos. Os casos extremos são os maiores ou menores que 2,5 vezes o quartil. Este tipo de gráfico permite observações quanto à frequência de valores, sua mediana e o intervalo de concentração de casos nos valores mais frequentes.

A análise da base de dados foi feita para visualizar a variação de resultados e suas tendências. Na avaliação de cada parâmetro, todos os casos são apresentados e estratificados para cada valor da variável. A Figura 12 apresenta os gráficos de carga anual de refrigeração estratificados pelas variáveis relacionadas à carga térmica interna.

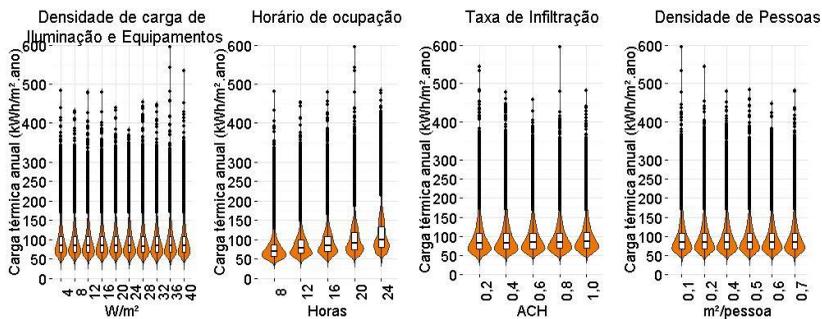


Figura 12. Influência dos parâmetros de carga interna na carga de refrigeração anual.

Comparando as medianas dos gráficos da Figura 12, além dos quartis e valores máximos e mínimos dos diagramas de caixa, é possível identificar os parâmetros cujos resultados de carga anual de refrigeração mais se diferenciam significativamente. Desconsiderando os valores extremos, observa-se que a densidade de pessoas, infiltração e carga interna pouco se diferenciam a partir dos valores variados. Diferente do efeito das horas de uso, que apresenta médias de cargas térmicas maiores conforme seus maiores valores. Porém, é importante salientar que para cada gráfico estão os resultados de todos os outros parâmetros combinados, estando estratificados pelos valores dos parâmetros.

A Figura 13 apresenta a influência dos parâmetros de parede na carga anual de refrigeração. Os parâmetros são de transmitância térmica, absorvância solar e capacidade térmica da parede.

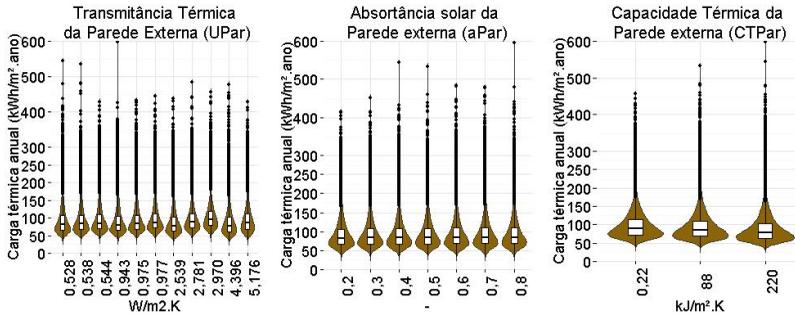


Figura 13. Influência dos parâmetros da parede externa na Carga de Refrigeração Anual.

Na Figura 13, os valores de absortância solar, através da análise das medianas e frequência dos resultados, não apresenta grupos de variância significativa. Por outro lado, para capacidade térmica, o aumento dos valores reduz a mediana, porém aumenta os valores extremos de carga anual de refrigeração. Há uma concentração maior de casos com menor carga térmica de refrigeração para os casos com capacidade térmica alta do que com os casos mais leves. Por sua vez, o parâmetro de transmitância térmica da parede apresenta uma não-linearidade em relação à carga anual de refrigeração.

A Figura 14 apresenta a influência dos parâmetros do telhado de carga anual de resfriamento, onde se observa um comportamento semelhante aos parâmetros de parede.

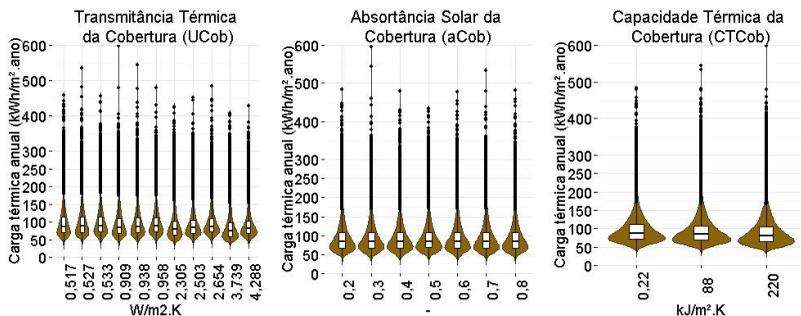


Figura 14. Influência dos parâmetros da cobertura na Carga de Refrigeração Anual.

Para os parâmetros de parede externa e cobertura não houve diferenças significativas para valores de cada variável. A grande combinação de parâmetros e a variabilidade da amostra permite que qualquer valor de parâmetro adotado para as paredes e coberturas possam resultar em qualquer valor de cargas térmicas de refrigeração, dependendo da combinação com as outras variáveis.

A influência da exposição da cobertura e piso na carga anual de refrigeração está apresentada na Figura 15.

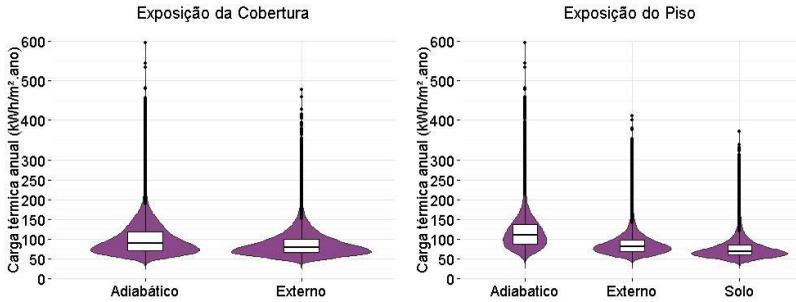


Figura 15. Influência dos parâmetros de exposição do piso e cobertura na Carga de Refrigeração Anual.

Na Figura 15, a exposição da cobertura para o exterior apresenta mais casos com carga de refrigeração anual com valores mais baixos que os casos em que a cobertura é adiabática. A cobertura adiabática possui mais casos com carga térmica maior que a mediana. Comparando as medianas e os quartis do diagrama de caixa da exposição da cobertura, pode-se dizer que existe uma diferença significativa entre os casos com cobertura voltada para o exterior e para as adiabáticas, e que as voltadas para o exterior geralmente apresentam valores menores. A exposição do piso demonstra que a maioria dos casos com o piso em contato com o solo apresenta carga anual de refrigeração menor do que os casos com o piso voltado para o exterior ou adiabático, sendo que os casos com piso adiabático têm mais da metade dos seus casos com valores maiores que o quartil superior dos casos com contato com o solo. As condições de exposição do piso diferenciam significativamente na carga térmica de refrigeração e geralmente os casos de contato com o solo têm cargas térmicas de refrigeração mais baixas que nas outras condições.

A Figura 16 mostra os “*violin plots*” da influência dos parâmetros da janela sobre a carga anual de refrigeração, apresentando os valores de

percentual de janela na fachada (WWR), transmitância térmica do vidro (U_{Vid}) e Fator Solar do Vidro (FS).

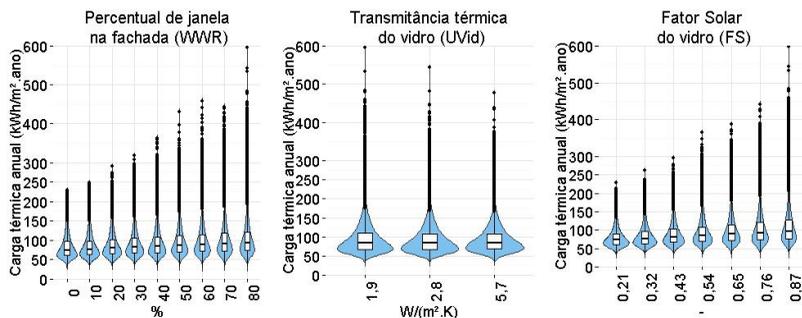


Figura 16. Influência dos parâmetros da janela na Carga de Refrigeração Anual.

Na Figura 16 é possível observar que quanto maior a área envidraçada da fachada e maior o fator solar dos vidros, maiores são as cargas anuais de refrigeração. Este resultado é esperado, pois o aumento destes valores acrescenta ganhos térmicos à zona por radiação solar. O aumento da transmitância térmica dos vidros apresentou pouca influência na carga anual de refrigeração da maioria dos casos. Sua maior influência foi apenas para os casos extremos, quando os casos com vidro simples, de maior transmitância térmica, apresentam menores valores extremos de carga térmica de refrigeração.

A influência dos ângulos de sombreamento sobre a carga anual de refrigeração é apresentado na Figura 17, com os ângulos verticais e horizontais de sombreamento (AVS e AHS) e ângulos de obstrução vizinha (AOV).

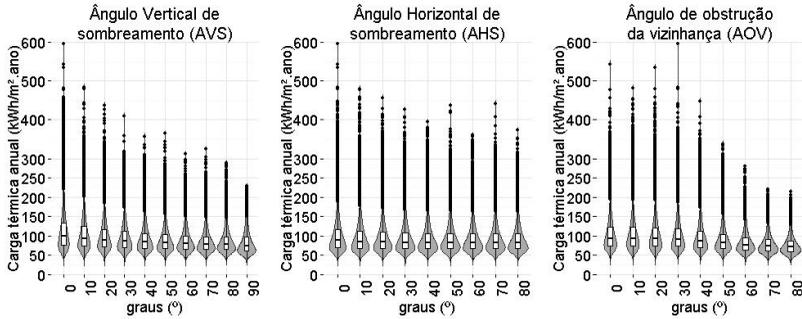


Figura 17. Influência dos parâmetros de sombreamento na Carga de Refrigeração Anual.

Conforme Figura 17, os ângulos verticais de sombreamento (AVS) apresentam uma grande influência nas cargas de refrigeração anual, pois quanto maior o ângulo menos a carga. Estes resultados ilustram como o sombreamento das aberturas influencia significativamente a redução de ganho de calor solar. Porém, para os ângulos horizontais de sombreamento (AHS), menor diferença é notada. Já o sombreamento causado pela vizinhança medido pelo ângulo de obstrução da vizinhança (AOV) apresenta significância com ângulos maiores de 40°. Considerando que à latitude 27 graus a altitude solar é cerca de 40° ao meio-dia no inverno e perto de 90° no verão, pode-se supor que o AOV não tem qualquer efeito significativo até atingir a elevação solar menor para as horas mais quentes do ano.

A Figura 18 apresenta os diagramas para orientação solar da fachada externa, altura do pé-direito da zona e área de massa da Inércia Térmica.

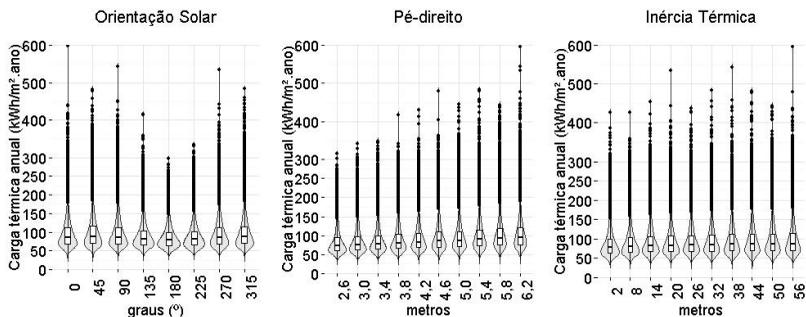


Figura 18. Influência dos parâmetros de orientação e pé-direito da zona térmica na Carga de Refrigeração Anual.

Na Figura 18, quanto à orientação solar, observa-se que todos os casos com valores de carga de refrigeração anual acima de $300 \text{ kWh/m}^2 \cdot \text{ano}$ não estão voltados para sul (180°). À nordeste (45°) e à noroeste (315°) apresentam os mais altos valores de cargas de refrigeração. Em relação ao pé-direito, os casos de carga de refrigeração anual acima de $300 \text{ Wh/m}^2 \cdot \text{ano}$ são maiores do que 2,6 m, sendo as medianas das cargas de refrigeração diretamente proporcionais ao pé-direito da zona térmica. A inércia térmica, representada pela área de massa de uma parede interna de concreto, apresenta pouca influência sobre a carga anual de refrigeração, com um leve aumento da carga de refrigeração anual com o acréscimo de inércia térmica.

Para todos os valores de parâmetros observados anteriormente se identifica cargas anuais de refrigeração próximas aos valores mínimos encontrados na amostra, cerca de $25 \text{ kWh/m}^2 \cdot \text{ano}$. Isto indica que com qualquer valor é possível ter resultados de carga de refrigeração baixa dependendo a combinação com os outros parâmetros. Mesmo com as maiores áreas envidraçadas na fachada ou maiores absorvência solares na cobertura, é sempre possível apresentar uma carga de refrigeração abaixo das médias.

No entanto, também é importante analisar a influência da combinação entre os parâmetros. Como exemplo deste cuidado, a Figura 19 ilustra a influência da área de janela na fachada em relação às diferentes orientações solares da fachada.

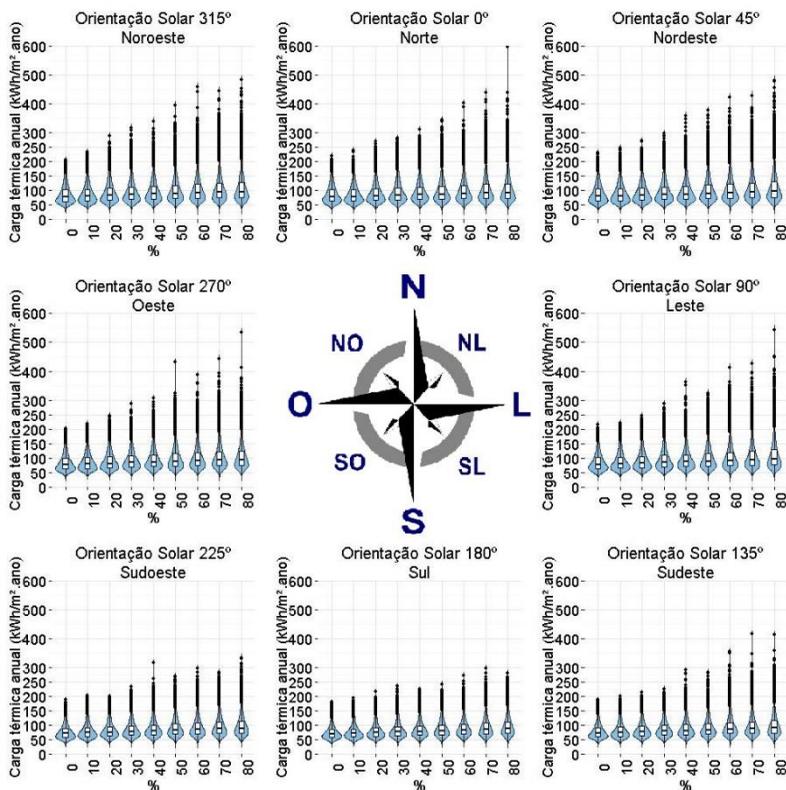


Figura 19. Influência do percentual de área de janela por Orientação Solar na Carga de Refrigeração Anual.

Conforme a Figura 19, pode-se observar que a carga de refrigeração dos casos com fachada orientada ao quadrante Sul, Sudoeste (135°), Sul (180°) e Sudeste (225°), sofrem menos influência da área de janela, principalmente para os valores menores que 40%. O mesmo ocorre

para os valores extremos, onde para orientação Sul, mesmo para 80% de área envidraçada, o maior valor não passa de 300kW/m².ano.

3.4. CONSIDERAÇÕES FINAIS

Neste capítulo foi construída uma base de dados com resultados de simulações de aproximadamente 1,29 milhões de casos. O objetivo desta base de dados é de representar um universo de combinações construtivas e de ocupação de uma zona térmica de edificação comercial condicionada artificialmente para elaboração de modelos estatísticos de predição de desempenho energético de edificações comerciais. Esta base de dados é a porção “caixa-branca” do modelo “caixa-cinza” que é proposta desta tese. Quanto à construção desta base de dados, destaca-se as seguintes considerações finais:

- a. Este processo de construção, através da parametrização, amostragem e simulações em cluster se mostrou como um método automático eficiente para construção de bases de dados, podendo ser reproduzido para outros climas ou outros parâmetros de interesse para análises de edificações;
- b. O desafio de visualizar e analisar o grande volume de dados foi facilitado com a adoção de diagramas como os “*violin plots*”, possibilitando a visualização de várias dimensões dos resultados em um único gráfico;
- c. A base de dados construída é considerada adequada para desenvolvimento de modelos de predição, pois há

variabilidade e homogeneidade nos parâmetros e valores adotados;

- d. Os resultados das simulações demonstram a inter-relação entre as variáveis, pois para qualquer valor de parâmetro existe um intervalo significativo de resultados de carga anual de refrigeração;
- e. Esta base de dados poderá ser utilizada para diversas outras investigações. O volume de informação pode ser utilizado para traçar tendências específicas de certos parâmetros ou combinações de parâmetros. Além disso, outras pesquisas poderão explorar os resultados de picos de refrigeração, e carga e pico de aquecimento.

4. DESENVOLVIMENTO E COMPARAÇÃO DE METAMODELOS

Neste capítulo a base de dados dos resultados das simulações elaboradas no capítulo anterior é utilizada para modelagens com seis técnicas de predição: Regressão linear múltipla, regressão adaptativa multivariada por splines, processo gaussiano, máquina de vetores de suporte, Randon forest e redes neurais artificiais. Para estas modelagens a base de dados foi processada para viabilizar a aplicação dos modelos. Todo processo estatístico foi feito em linguagem R (R, 2014) com as funcionalidades de desenvolvimento de modelos de predição do pacote Caret (KUHNS, 2015). Por fim, os modelos foram comparados pelo tempo de treinamento, coeficiente de determinação de Spearman, raiz do erro médio quadrático e raiz do erro médio quadrático normalizada.

A Figura 20 ilustra o método de comparação dos metamodelos. A partir do universo de mais de cinco quatrilhões de combinações possíveis entre os parâmetros adotados foi amostrado 1,29 milhões de casos. Destes casos simulados, foi selecionada uma amostra para treinamento dos seis metamodelos, que foram utilizados para predição dos resultados dos mesmos 1,29 milhões de casos simulados. Ao final, a comparação entre os resultados simulados e estimados pelo metamodelo foi utilizada para comparar seus desempenhos.

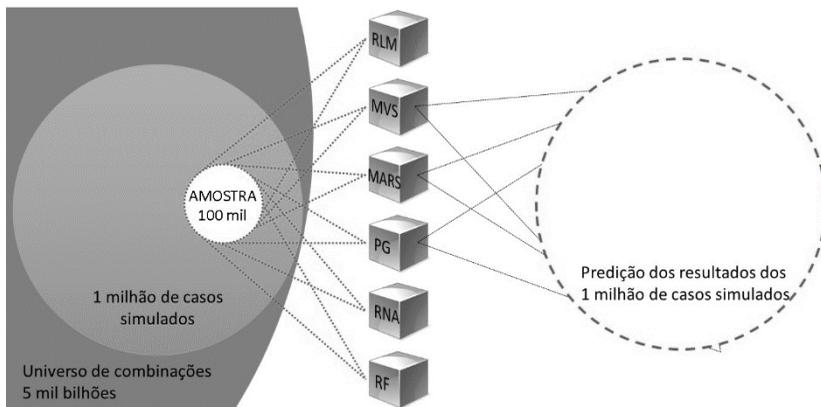


Figura 20. Esquema do método de comparação dos metamodelos

4.1. MÉTODO: PRÉ-PROCESSAMENTO

Antes de preparar os metamodelos é importante garantir que os dados satisfaçam algumas premissas para eficiência no seu desenvolvimento. Como pré-processamento dos dados, foi avaliada a possibilidade de exclusão de parâmetros desnecessários, a adição de parâmetros substitutos e a transformação das unidades para normalização dos resultados. Além disso, devido ao fato dos parâmetros não terem a mesma unidade e grandeza, os dados foram centralizados e escalonados.

4.1.1. EXCLUSÃO DE PARÂMETROS CORRELACIONADOS

Para analisar a correlação entre os parâmetros foi elaborada uma matriz de correlação entre os valores de todas as variáveis. A Figura 21 apresenta a matriz de correlação onde é apresentado um gráfico polar para cada par correlacionado. É feita a correlação de cada parâmetro com todos os outros, sendo que o tamanho da fatia do gráfico de pizza e sua cor

representam a magnitude de correlação entre os parâmetros. A cor azul representa correlação positiva e a vermelha, negativa.

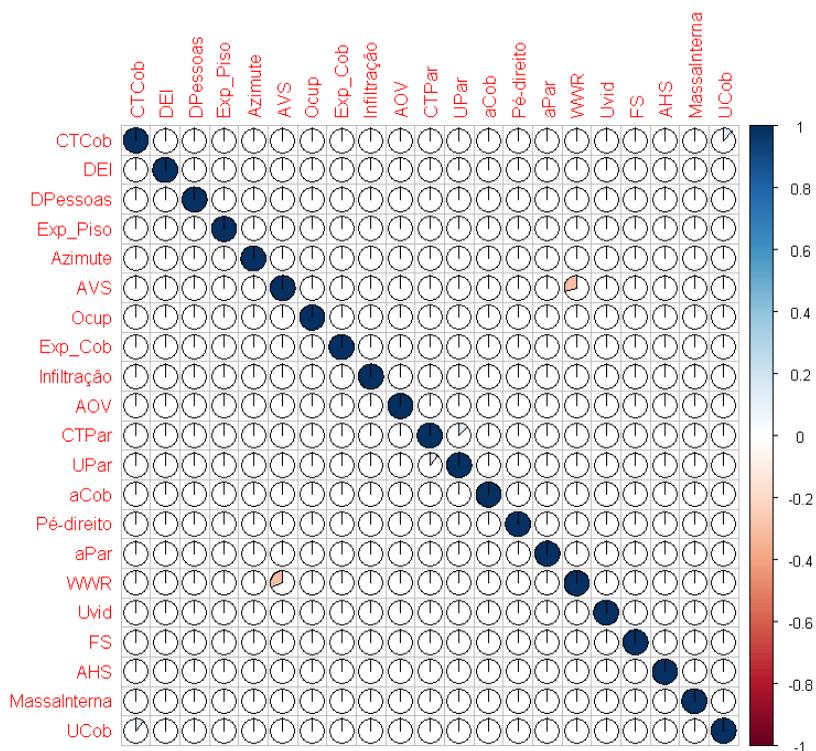


Figura 21. Matriz de correlação entre as variáveis preditoras

A matriz de correlação mostra que apenas algumas combinações apresentam correlações, porém estas são menores que 0,5. O que permite concluir que todos os parâmetros são independentes uns dos outros e podem ser utilizados na modelagem. Esta característica é importante para evitar parâmetros redundantes, o que dá vantagens matemáticas no desenvolvimento dos modelos.

4.1.2. ADIÇÃO DE VARIÁVEIS SUBSTITUTAS

Quando há um parâmetro de categoria, como parâmetros que definem tipos ou condições que não têm representação numérica, é recomendado que sejam decompostos em variáveis mais específicas (KUHN; JOHNSON, 2013). Por isso, os parâmetros de exposição da cobertura e exposição do piso foram transformados em variáveis binárias substitutas. A Tabela 5 apresenta as variáveis substitutas utilizadas nas modelagens estatísticas.

Tabela 5. Variáveis substitutas

Parâmetro	Unidade	Valores
Exposição da cobertura	Binária	0-adiabático; 1-voltado para o exterior.
Piso em contato com o solo	Binária	0-não; 1-sim
Piso voltado para o exterior	Binária	0-não; 1-sim
Piso adiabático	Binária	0-não; 1-sim

O parâmetro de exposição da cobertura foi transformado em uma variável binária onde 1 (um) representa a exposição da cobertura para o exterior e 0 (zero) representa a cobertura adiabática. Já o parâmetro de exposição do piso teve que ser decomposto em três variáveis: piso adiabático; piso em contato com o solo; e piso voltado para o exterior. Estas três variáveis são binárias, onde 1 (um) representa a existência da condição e 0(zero) a não existência. Por exemplo, se a variável Piso adiabático é 1 significa que o piso é adiabático, porém se a variável for 0(zero) o piso pode estar exposto ao exterior ou em contato com o solo. Sendo esta segunda condição definida pelas outras variáveis substitutas.

4.1.3. TRANSFORMAÇÃO DA VARIÁVEL DEPENDENTE

A análise dos dados de entrada para elaboração de modelos estatísticos requer uma atenção especial para garantir a normalidade de ocorrência dos valores de cada parâmetro. As variáveis independentes, parâmetros preditores, foram modeladas de modo a terem distribuição homogênea, não requerendo ajustes. Entretanto, os valores da variável dependente, carga térmica de refrigeração, são resultados das simulações, por isso foram analisadas quanto a sua distribuição. Uma distribuição não enviesada para nenhum lado é uma distribuição simétrica. Isso significa que a probabilidade de ocorrência para qualquer lado da média dos resultados é igual (KUHNS and JOHNSON, 2013). Foi utilizado o método de transformação box-cox para testar o enviesamento dos resultados das simulações e definir qual a melhor forma de transformação apropriada para corrigir esse desvio. O método box-cox corresponde a uma família de funções estatísticas para criar uma classificação a partir de um parâmetro denominado λ .

$$x^* = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(x) & \text{se } \lambda = 0 \end{cases} \quad (4)$$

O valor de λ é obtido a partir dos dados do parâmetro testado com as funções de box-cox. Conforme a Equação 4, a transformação deve ser logarítmica se a família de funções resultar em um $\lambda=0$. Porém, ainda em função de λ , é possível identificar a necessidade de transformações por quadrado ($\lambda=2$), raiz quadrada ($\lambda=0,5$), inverso ($\lambda=-1$) e outras transformações intermediárias. Esta transformação foi aplicada na variável de carga térmica anual de refrigeração através da linguagem R,

utilizando a função box-cox do pacote MASS (VENABLES; RIPLEY, 2002).

A avaliação da distribuição da variável de carga térmica anual de refrigeração indicou uma transformação logarítmica para resolver o enviesamento à esquerda apresentado por esta variável. A Figura 22 apresenta a distribuição da carga de refrigeração antes e depois da transformação.

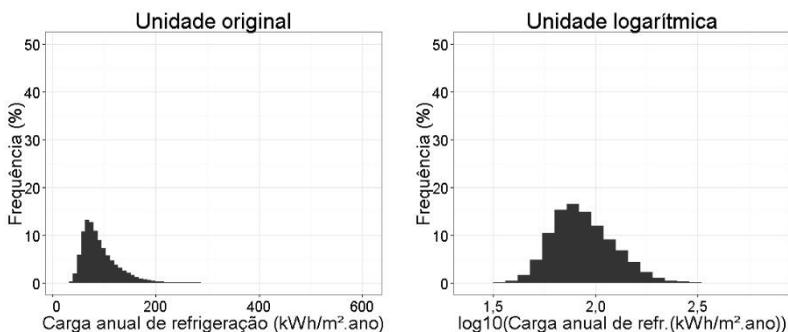


Figura 22. Distribuição das Cargas Anuais de Refrigeração antes e depois da transformação das unidades

A distribuição com as unidades transformadas não são perfeitamente simétricas com a transformação logarítmica, mas estão mais centralizadas do que considerando as unidades originais.

4.1.4. CENTRALIZAÇÃO E ESCALONAMENTO

A maioria das técnicas de modelagem têm requisitos rigorosos para os dados, sendo que os preditores devem estar em uma escala comum. Como os preditores da base de dados deste trabalho têm diferentes unidades e magnitudes de valores, os dados foram centralizados e escalonados para melhorar a estabilidade numérica dos cálculos. Para

centralizar os dados, o valor médio das variáveis preditoras foi subtraído de todos seus valores, induzindo as variáveis preditoras a ter uma média zero. Para o escalonamento, os valores das variáveis preditoras foram divididos por seus desvios padrões, transformando os valores para que as variáveis preditoras tenham um desvio padrão comum. Este processo é feito pelas funções *center* e *scale*, próprias para pré-processamento pelo pacote *caret* (KUHN, 2015). Após o processamento dos metamodelos, as variáveis são transformadas de volta, para se obter os resultados na unidade original.

4.2. MÉTODO: TREINAMENTO DOS METAMODELOS

Seis técnicas de predição foram utilizadas para o desenvolvimento dos modelos: regressão linear múltipla, regressão adaptativa multivariada por splines, processo gaussiano, máquina de vetores de suporte, Random forest e redes neurais artificiais. Os modelos foram treinados com uma amostra dimensionada para garantir uma confiança satisfatória. Foi adotado um método de validação cruzada para garantir que a amostra e suas configurações fossem aceitáveis.

Para o treinamento dos modelos foi montada uma máquina virtual em um servidor computacional configurada para o processamento em R. A Máquina Virtual foi montada com 8 núcleos de 3,33GHz de processamento. Como o R trabalha com o processamento de dados totalmente carregados na memória RAM, foram disponibilizados 16GB de memória para garantir que toda base de dados pudesse ser processada.

O desenvolvimento de todos os modelos foi realizado com o pacote *caret* para R (KUHN, 2015). O pacote possui ferramentas para tratamento

de dados, pré-processamento de variáveis, seleção de configurações dos modelos e avaliação de importância de variáveis. Porém, sua maior funcionalidade é padronizar a linguagem de diversos pacotes estatísticos de regressão, permitindo ser elaborados modelos similares desenvolvidos por diferentes técnicas estatísticas. Os modelos foram realizados com suas configurações padrão. A regressão linear múltipla foi elaborada com o pacote ‘*e1071*’ (MEYER et al., 2014). Para o modelo de redes neurais artificiais foi selecionado o pacote ‘*NNET*’ (RIPLEY; VENABLES, 2014). O modelo de regressão multivariada por splines foi desenvolvido com o pacote ‘*Earth*’ (MILBORROW, 2015). Para o modelo de máquina de vetores de suporte foi selecionado o pacote ‘*svmLinear*’ (MEYER et al., 2014). Para o treinamento do modelo Random Forest foi usado o pacote ‘*randomForest*’ (BREIMAN AND CUTLER, 2014). E para o modelo de processo gaussiano foi usado o pacote ‘*gaussprLinear*’ (KARATZOGLOU; SMOLA; HORNIK, 2015).

4.2.1. DIMENSÃO DA AMOSTRA PARA TREINO

O dimensionamento de uma amostra para elaboração de modelos é definido para se obter o menor custo de aquisição de dados com a garantia de confiança da amostra. Para se definir uma amostra que permita calcular estimativas suficientemente precisas dos resultados, sua dimensão depende da variabilidade do universo e sua complexidade. Existem técnicas que permitem obter um número mínimo de casos para uma amostra, garantindo uma determinada precisão. Com essa garantia, escolher uma amostra maior é questão de considerar os custos de desenvolvimento e os ganhos de precisão.

A amostra para treinamentos foi dimensionada através da estimativa pela média populacional. A Equação 5 é dada para se calcular o tamanho da amostra em relação a uma estimativa confiável da média populacional:

$$n = \frac{N \cdot \sigma^2 \cdot (Z_{\alpha/2})^2}{(N - 1) \cdot E^2 + \sigma^2 \cdot (Z_{\alpha/2})^2} \quad (5)$$

Onde:

n é o número de casos da amostra;

N é o número de casos do universo;

σ é o desvio padrão do universo da variável estudada (kWh/m².ano);

E é o erro máximo desejado da estimativa. Na mesma unidade da variável;

$Z_{\alpha/2}$ é o Valor crítico correspondente ao grau de confiança desejado, conforme Tabela 6.

Tabela 6. Exemplo de críticos associados ao grau de confiança da amostra

Intervalo de confiança	Valor Crítico $Z_{\alpha/2}$
90%	1,645
95%	1,96
99%	2,575
99,5%	3,090
99,9%	3,291

Este cálculo foi feito para a carga anual de refrigeração conforme os resultados das simulações: desvio padrão de 28,7 kW/m².ano para uma população de 1.293.250 casos. Considerando um erro máximo da estimativa em 1% do valor do desvio padrão, ou seja, 0,287kW/m².ano de erro máximo desejado, e um nível de confiança esperado de 99,9%, a amostra foi dimensionada em 100.512 casos. Para os treinamentos, o tamanho da amostra foi arredondado para 100 mil casos selecionados pelo método de Hipercubo Latino.

4.2.2. VALIDAÇÃO CRUZADA

Para validar o treinamento dos modelos foi utilizada uma técnica de validação cruzada (*cross-validation*) que garante o desempenho do modelo comparando a similaridade de vários treinamentos realizados com dados reamostrados da amostra de treino. Uma subamostra dos dados é utilizada para treinar um modelo e o restante dos dados é usado para avaliar sua eficácia. Este processo é repetido múltiplas vezes e os resultados são estudados para verificar se os modelos treinados são similares. A similaridade dos modelos valida a hipótese de que pode-se representar o comportamento do sistema a partir das configurações da modelagem testada (KUHN; JOHNSON, 2013).

Duas técnicas de validação cruzada foram utilizadas. Primeiramente, uma amostra de 75% dos casos para treinamento foi separada para treino e o restante guardado para teste. Este processo foi repetido oito vezes, ou seja, oito amostras aleatórias de 75% dos 100 mil casos foram usadas. Para cada amostra destas foi aplicada a técnica de validação cruzada de k-fold, onde amostras são aleatoriamente particionadas em k conjuntos aproximadamente de mesmo tamanho. Um modelo é treinado usando todas as subamostras, exceto a primeira (chamada de primeiro fold). Então o modelo é utilizado para predizer os dados da primeira amostra e medir seu desempenho. Este treino é repetido para cada partição, sempre utilizando os dados remanescentes para teste. Estes processo é ilustrado pelo esquema da Figura 23.

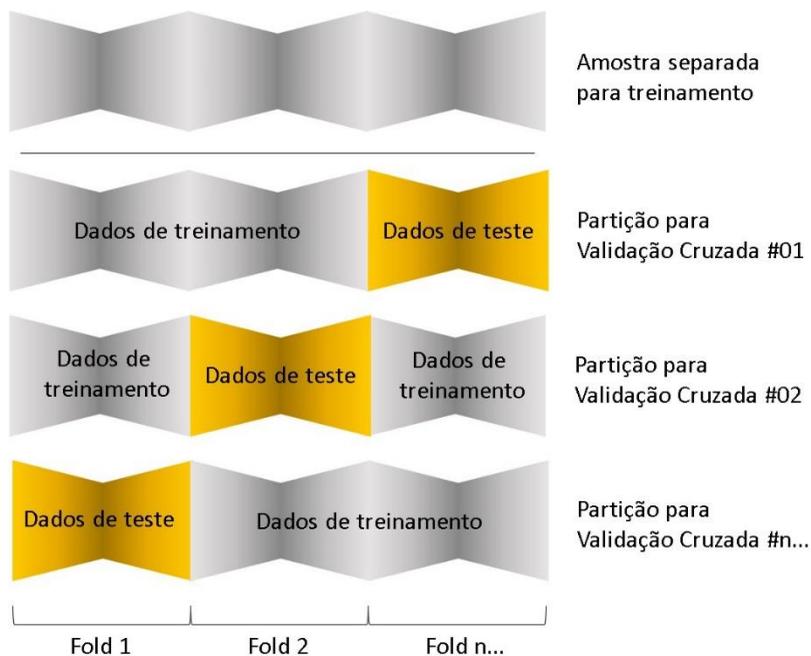


Figura 23. Esquema de validação cruzada de 3 folds.

Os seis modelos estatísticos foram processados com 8-folds para validação cruzada e cada rodada de validação foi repetida oito vezes. Portanto, cada modelo foi treinado 64 vezes. A variação dos resultados de R^2 e RMSE de todos os treinamentos da validação cruzada foi avaliada e, sendo satisfatória, o melhor modelo treinado foi adotado para reproduzir a base de dados completa, de 1,29 milhões de casos.

4.3. MÉTODO: AVALIAÇÃO E COMPARAÇÃO DOS METAMODELOS

Para avaliação e comparação dos metamodelos foram adotados quatro índices de desempenho: tempo de treinamento; coeficiente de determinação de Spearman (R^2); raiz do erro médio quadrático (RMSE) e

erro médio quadrático normalizado (NRMSE). O tempo foi registrado durante o treinamento dos modelos e o R^2 , RMSE e NRMSE foram calculados a partir da aplicação dos modelos para predição de toda base de dados, ou seja, para predição dos 1,29 milhões de casos. Estes critérios de desempenho foram combinados em um índice de desempenho global (IDG) para se determinar qual modelo apresentou o melhor desempenho.

O coeficiente de determinação de Spearman (R^2) foi usado para medir o grau de correlação das variáveis dependentes entre os valores simulados e preditos pelos modelos. Este coeficiente também dá informações quanto à direção da correlação, positiva ou negativa. Foi adotado o coeficiente de Spearman porque, diferente do coeficiente de determinação de Pearson, não é uma premissa que os valores simulados e preditos tenham uma correlação linear.

A raiz do erro médio quadrático é a raiz quadrada da média quadrática dos erros, o quadrado das somas dos resíduos entre os casos simulados e os preditos. Este valor é interpretado como o quão longe os resíduos estão de zero ou da distância média entre os valores simulados e as predições do modelo. Esta análise foi traçada em um gráfico de dispersão dos valores das Cargas Anuais de Refrigeração simulados versus os valores preditos pelo modelo. O valor do RMSE é dado na mesma unidade das variáveis que avalia, ou seja, kWh/m².ano.

A raiz do erro médio quadrático normalizada (NRMSE) é o valor normalizado do RMSE, sendo representado em percentual. Representa o percentual do erro médio quadrático entre os resultados máximos e mínimos encontrados na base de dados.

Foi calculado o índice de síntese de desempenho global baseado no índice de síntese (SI) de Chou e Bui (2014). Este índice é calculado pelo

somatório da média da soma das razões do desempenho de cada modelo em comparação com os modelos de maior e menor desempenho, conforme Equação 6.

$$IDG = \frac{1}{n} \sum_{i=1}^n \left(\frac{P_i - P_{min,i}}{P_{max,i} - P_{min,i}} \right) \quad (6)$$

Onde:

n é o número de critérios de desempenho;

P_i é o indicador de desempenho adotado.

Como critérios de desempenho foram considerados o tempo de treinamento, o R^2 e o RMSE. Este cálculo considera os maiores valores dos critérios como indicadores dos melhores desempenhos, entretanto o tempo e o RMSE são critérios dos quais quanto menor o valor, melhor seu desempenho. Logo, os valores de tempo e RMSE foram invertidos para aplicação do cálculo de desempenho global, ou seja, foram transformados em valores negativos para que o menor valor se torne o maior. O maior valor de IDG indica uma alta precisão do modelo de predição.

4.4. RESULTADOS: METAMODELOS

A Figura 24 apresenta o tempo de treinamento de cada Metamodelo em horas. É possível observar que cada técnica requer esforços de processos computacionais diferente.

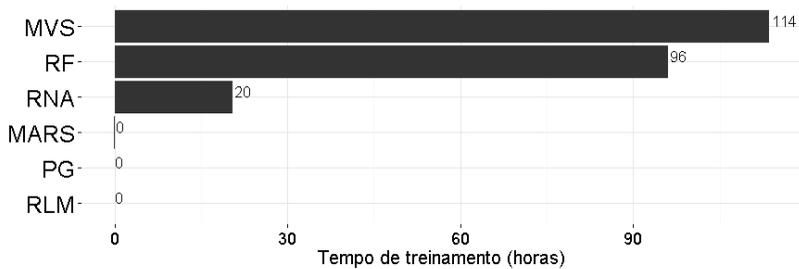


Figura 24. Tempo de treinamento dos metamodelos

As máquinas de vetor de suporte (MVS) e Random Forest (RF) foram as técnicas que exigiram mais horas para treinamento, 113,6 e 96,1 horas, respectivamente. As redes neurais artificiais levaram 20,5 horas, e as demais técnicas levaram menos de 10 minutos cada.

Para o processo de validação cruzada cada modelo foi treinado com oito repetições de amostras aleatórias com 75% dos 100 mil casos de treinamento, cada repetição foi treinada com oito reamostragens da validação cruzada k-fold. A Tabela 7 apresenta o resumo dos resultados do processo de validação cruzada com os valores de coeficiente de determinação de Spearman (R^2) e a raiz do erro médio quadrático (RMSE). Estes valores são resultados da predição de cada treinamento para as amostras de validação.

Tabela 7. R² e RMSE dos treinamentos de validação cruzada

	Spearman R ²			RMSE (kWh/m ² .ano)		
	Mín.	Média	Máx.	Mín.	Média	Máx.
MVS	0,750	0,756	0,762	7,31	7,38	7,46
RLM	0,756	0,760	0,763	7,23	7,29	7,40
MARS	0,789	0,794	0,799	6,66	6,74	6,81
RF	0,925	0,927	0,929	4,09	4,13	4,21
PG	0,970	0,971	0,972	2,50	2,55	2,58
RNA	0,988	0,990	0,991	1,42	1,48	1,62

Conforme a Tabela 7, o processo de validação cruzada demonstrou que o tamanho da amostra e as variáveis foram suficientes para estabilizar os treinamentos. Os seis modelos apresentaram pouca variação entre os valores máximos e mínimos de R² e RMSE. As variações de R² foram menores que 2%, sendo a maior diferença apresentada pela Máquina de Vetores de Suporte (MVS), com 1,6%. Para a raiz do erro médio quadrático, RMSE, a maior variação relativa foi do metamodelo de Rede Neural Artificial (RNA), porém este mesmo modelo apresentou a menor variação absoluta. Para todos os metamodelos não houve variações significativas, sendo todos considerados válidos. O modelo de melhor desempenho da validação cruzada foi adotado como metamodelo final, e aplicado para predição de toda base de dados.

As próximas análises usaram os metamodelos treinados com cem mil casos para predição de toda base de dados, ou seja, 1,29 milhões de casos. A Tabela 8 apresenta as médias e desvios padrão da população de resultados da simulação e dos resultados preditos pelos metamodelos.

Tabela 8. Média e desvio padrão dos resultados simulados e preditos pelos metamodelos

	Média (kWh/m².ano)	Desvio padrão (kWh/m².ano)	Diferença da média para o simulado (%)	Diferença do desvio padrão para o simulado (%)
Simulado	75,09092	28,69551		
RLM	73,86162	22,57699	-2	-21
MARS	74,02129	23,20734	-1	-19
PG	74,83387	27,53730	0	-4
MVS	72,53243	21,75001	-3	-24
RNA	75,03599	28,53775	0	-1
RF	74,22182	25,02869	-1	-13

Conforme a Tabela 8, os metamodelos conseguiram reproduzir a média da população dos casos simulados com diferença de até 3% e com desvio padrão até 21% menores que o desvio padrão apresentado pela população de resultados simulados. Os metamodelos representam bem a média, mas possuem desvios padrões menores que a população real (simulada), isto indica que os resultados dos metamodelos tendem a ser mais próximos da média. A reprodução da média e desvio padrão só dizem respeito a reprodução da população. Como o interesse dos metamodelos está na predição individual de cada caso, os valores de coeficiente de determinação e a raiz do erro médio quadrático podem ilustrar melhor este desempenho.

As correlações entre os resultados previstos e simulados de carga anual de refrigeração para cada Metamodelo são apresentadas na Figura 25. Além disso, nos gráficos há os resultados do coeficiente de determinação (R^2), raiz do erro médio quadrático (RMSE) e o RMSE normalizado (NRMSE). Os gráficos de dispersão de valores preditos e simulados são apresentados com a densidade de ocorrência pela concentração de pontos. A concentração de pontos é realizada pela cor mais clara.

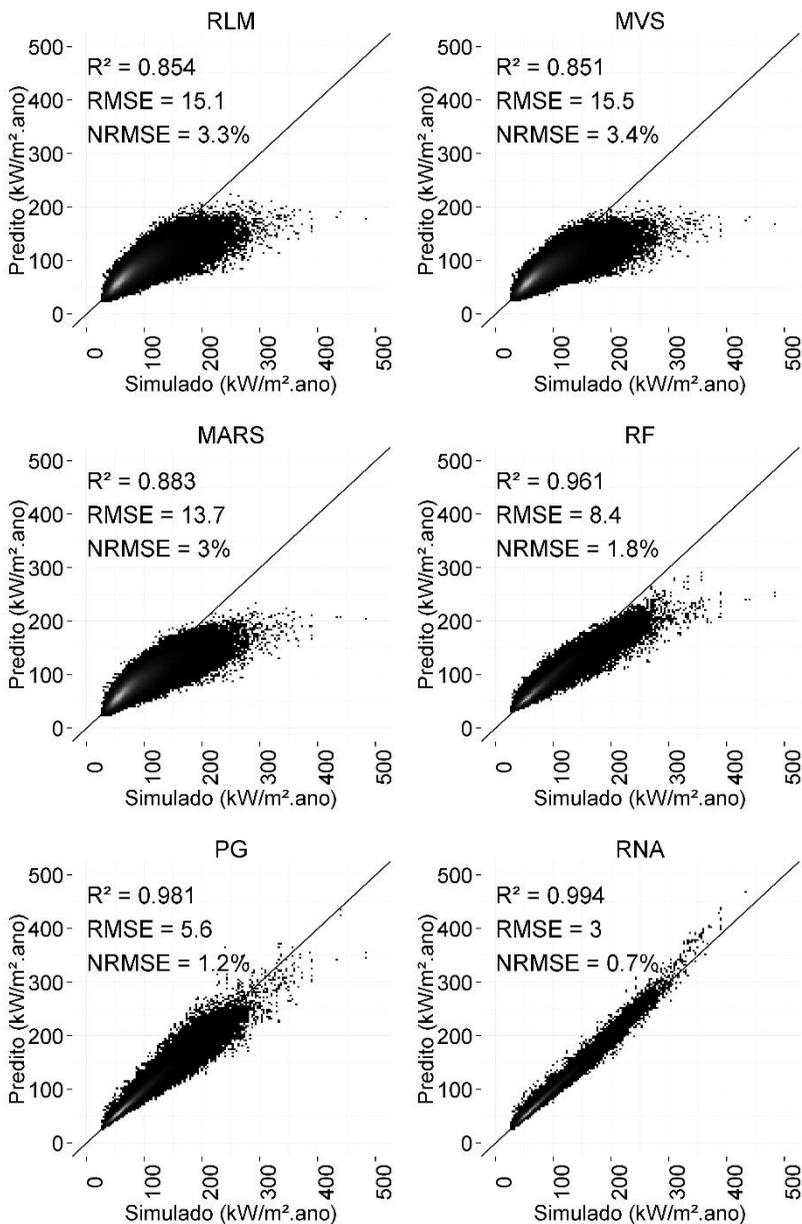


Figura 25. Gráficos de dispersão e índices de desempenho dos metamodelos

A linha de tendência diagonal nos gráficos de dispersão demonstra uma tendência linear entre os resultados previstos e simulados dos seis metamodelos. No entanto nos metamodelos de RLM, MVS e MARS existem pontos com pouco grau de associação. Estes casos formam uma nuvem de pontos abaixo da linha diagonal, o que corresponde a uma associação negativa entre os resultados previstos e simulados. Esta nuvem de pontos tem uma influência negativa sobre o R^2 .

Os valores de R^2 e NRMSE dos metamodelos aplicados para toda base são diferentes dos resultados dos treinamentos apresentados na Tabela 7, de validação cruzada. Esta diferença é causada pelo maior número de casos considerados na avaliação da base de dados completa, onde há mais possibilidade de correlação.

Por mais que o pré-processamento tenha reduzido o enviezamento dos valores de carga anual de refrigeração, alguns modelos ainda foram influenciados pelo deslocamento não resolvido pela transformação logarítmica da unidade da variável dependente. Esta influência fica clara observando os pontos com grau de associação negativo.

Os metamodelos RNA e PG apresentaram os maiores valores de R^2 , 0,994 e 0,981, respectivamente. Estes modelos apresentam a melhor dispersão linear e, conseqüentemente, a menor raiz do erro médio quadrático (RMSE). O Metamodelo de Redes Neurais Artificiais obteve um percentual de erro de 0,7%, o menor comparado com os resultados NRMSE dos outros modelos.

Carlo e Lamberts (2008) e Melo (2012) desenvolveram metamodelos que são considerados precedentes a esta tese. Ambos elaboraram metamodelos para edificações comerciais para o clima de Florianópolis baseados em simulações com o EnergyPlus. O primeiro

com metamodelos de regressão linear múltipla e o segundo com redes neurais artificiais. Apesar de não serem comparáveis com os metamodelos desenvolvidos aqui, por predizerem consumo de energia elétrica de uma edificação completa e não a carga térmica de refrigeração de uma zona térmica, seus resultados são referência para a análise de desempenho dos metamodelos aqui apresentados.

O Metamodelo de Carlo e Lamberts (2008) alcançou um RMSE de 8,06kWh/m².ano com R² de 0,9824. Como evolução deste resultado, Melo (2012) alcançou RMSE de 3,7kWh/m².ano com R² de 0,97. Apesar de não representarem os desempenhos frente ao mesmo número de casos e a mesma variabilidade de parâmetros, este trabalho alcançou RMSE de 3kWh/m².ano e R² de 0,994 com o metamodelo de rede neural artificial, com melhor desempenho que seus precedentes.

A distribuição de probabilidade de erros reais para cada Metamodelo é apresentada na Figura 26. Esta distribuição permite observar a concentração dos desvios observados nas dispersões dos valores de carga anual de refrigeração simulados versus preditos.

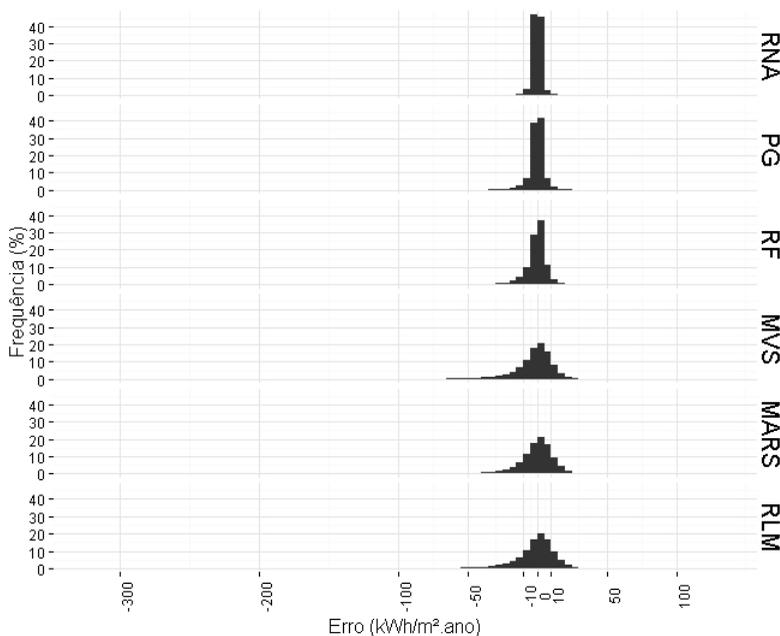


Figura 26. Distribuição de erros reais

Na Figura 26 observa-se que os erros de todos os metamodelos apresentam uma distribuição quase simétrica, tendo maior concentração de erros próximos a zero. Esta distribuição está relacionada com a precisão de desempenho do metamodelo. É importante mencionar que o eixo zero está deslocado para a direita porque alguns dos metamodelos apresentaram erros negativos. A distribuição mostra erros de até 300kWh/m².ano, mas a concentração dos erros é próxima de zero para todos os metamodelos.

O Índice de Desempenho Global (IDG) foi calculado com o tempo de treinamento dos modelos e o R² e o RMSE resultante da aplicação dos modelos na predição da base de dados completa. A Tabela 9 apresenta os valores de IDG. Os valores de Tempo e RMSE foram invertidos para

valores negativos a fim de orientar o maior desempenho para os maiores valores. Além disso, o valor de IDG foi normalizado para comparação dos valores entre 0 e 1.

Tabela 9. Índices de desempenho global dos metamodelos

	Tempo.(-1) horas	R ²	RMSE.(-1) kWh/m ² .an o	IDG	IDG Normal.
MVS	-113,573	85,1	-15,5	-0,83	0,00
PG	-0,010	98,1	-5,6	0,07	0,96
RNA	-20,468	99,4	-3,0	0,11	1,00
RLM	-0,008	85,4	-15,1	-0,48	0,37
MARS	-0,088	88,3	-13,7	-0,37	0,48
RF	-96,126	96,1	-8,4	-0,33	0,53

O metamodelo de Máquina de Vetores de Suporte obteve o pior desempenho, seguido pela Regressão Linear Múltipla. A diferença de desempenho de R² e RMSE destes dois modelos é baixa, porém o tempo de treinamento da Máquina de Vetores de Suporte foi o maior entre todos os metamodelos, o que exaltou sua pior colocação. Os melhores metamodelos, pelo índice de desempenho global, foram o de Rede Neural Artificial e de Processo Gaussiano. E mesmo com a Rede Neural Artificial levando vinte horas a mais de treinamento do que o processo Gaussiano, a Rede Neural Artificial ainda obteve o melhor desempenho.

4.5. CONSIDERAÇÕES FINAIS

Neste capítulo foram desenvolvidos seis metamodelos através de técnicas estatísticas diferentes. Os dados da base construída no capítulo anterior foram pré-processados e uma amostra de cem mil casos foi separada para o desenvolvimento dos metamodelos. Os metamodelos

foram treinados e comparados, permitindo as seguintes considerações finais:

- a. Os treinamentos dos modelos demandaram tempos muito diferentes. O entendimento desta variável é importante para elaboração de futuros trabalhos para outros climas;
- b. O processo de validação cruzada demonstrou que o tamanho da amostra e as variáveis foram suficientes para estabilizar os treinamentos;
- c. As seis técnicas estatísticas de predição apresentaram tendência linear entre os resultados simulados e preditos, apresentando médias equivalentes aos casos simulados e desvios padrões até 25% menores;
- d. O metamodelo de Rede Neural Artificial apresentou o melhor desempenho entre os metamodelos treinados, portanto foi selecionado para ser aperfeiçoado com o estudo paramétrico de suas configurações principais;
- e. Os outros metamodelos poderiam mudar seu desempenho através da configuração em seu desenvolvimento, porém aqui foram consideradas as propriedades padrões dos pacotes estatísticos.

5. REFINAMENTO DO METAMODELO DE REDE NEURAL ARTIFICIAL

O Metamodelo de Rede Neural Artificial foi escolhido como o que melhor representou a natureza do comportamento da zona térmica. Entretanto, mesmo apresentando índices de desempenho satisfatório, buscou-se aperfeiçoar este modelo através de uma análise paramétrica de seus principais parâmetros de modelagem. Para isso, foi realizado um refinamento da arquitetura da Rede Neural Artificial para definir o tamanho de amostra e tamanho da camada interna. Foram adotados 15 tamanhos de amostras e 4 tamanhos de camada interna, num total de 60 treinamentos de Redes Neurais Artificiais.

A Figura 27 ilustra o método, diferentes tamanhos de amostras são selecionados a partir de um milhão de casos simulados para elaboração de metamodelos com variação da estrutura de camadas internas. Estes metamodelos são utilizados para estimar os resultados de um milhão de casos simulados a fim de serem comparados quanto ao seu desempenho em coeficiente de determinação, raiz do erro médio quadrático e tempo de processamento.

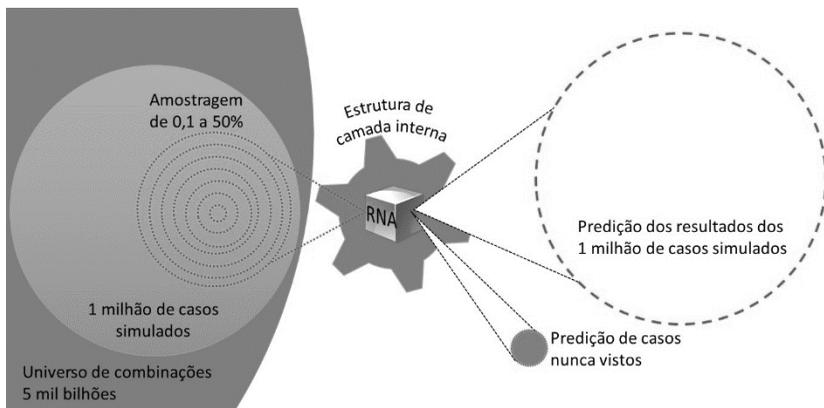


Figura 27. Esquema do método de refinamento da Rede Neural Artificial.

5.1. MÉTODO: PARÂMETROS DAS ARQUITETURAS DE RNAS

Todos os modelos treinados para este refinamento foram elaborados com o mesmo pré-processamento dos dados e validação cruzada apresentada no capítulo 4. A mesma máquina virtual configurada para o processamento em R foi utilizada, onde as repetições de validação cruzada de cada rede neural artificial rodaram em paralelo nos 8 núcleos de 3,33GHz de processamento. Além disso, 16GB de memória RAM foram disponibilizados para garantir que toda a complexidade das Redes Neurais Artificiais fosse processada.

5.1.1. TAMANHO DA AMOSTRA

O tamanho da amostra separada para treinamento do modelo está relacionada ao seu intervalo de confiança e ao erro máximo estimado, e consequentemente a precisão do modelo de predição. Considerando a população de aproximadamente 1,29 milhões de casos foram

determinados três grupos de tamanhos de amostras para treinamento, um grupo com amostras de 0,1, 0,2, 0,3, 0,4 e 0,5% da população simulada, outro com 1, 2, 3, 4 e 5% e o terceiro com 10, 20, 30, 40 e 50% da população. O tamanho destes grupos de amostras foram definidos arbitrariamente. Estes quinze valores foram utilizados como a variável de tamanho de amostra para refinamento do modelo de Rede Neural Artificial.

A equação de dimensionamento de amostra pela média populacional (Equação 2) foi utilizada para calcular os erros máximos estimados de cada tamanho de amostra. A partir do desvio padrão de 28,9kWh/m².ano e o intervalo de confiança de 99,9% foram calculados os erros máximos estimados apresentados na Tabela 10.

Tabela 10. Erros máximos estimados para diferentes tamanhos de amostra.

Percentual da População	Tamanho da Amostra	Erro máximo estimado (kWh/m².ano)	Erro máximo estimado para o desvio padrão (%)
0,1%	1.293	2,652	9,15%
0,2%	2.586	1,874	6,47%
0,3%	3.880	1,529	5,28%
0,4%	5.173	1,324	4,57%
0,5%	6.466	1,324	4,57%
1,0%	12.933	0,835	2,88%
2,0%	25.865	0,587	2,03%
3,0%	38.798	0,477	1,65%
4,0%	51.730	0,411	1,42%
5,0%	64.663	0,366	1,26%
10,0%	129.325	0,252	0,87%
20,0%	258.650	0,168	0,58%
30,0%	387.975	0,128	0,44%
40,0%	517.300	0,103	0,35%
50,0%	646.625	0,084	0,29%

Os erros máximos estimados pelas amostras variaram de 2,652 a 0,084kWh/m².ano. Onde quanto maior o tamanho da amostra, menor é

seu erro estimado. A Figura 28 apresenta o gráfico de dispersão entre o percentual da população e erro máximo estimado das amostras.

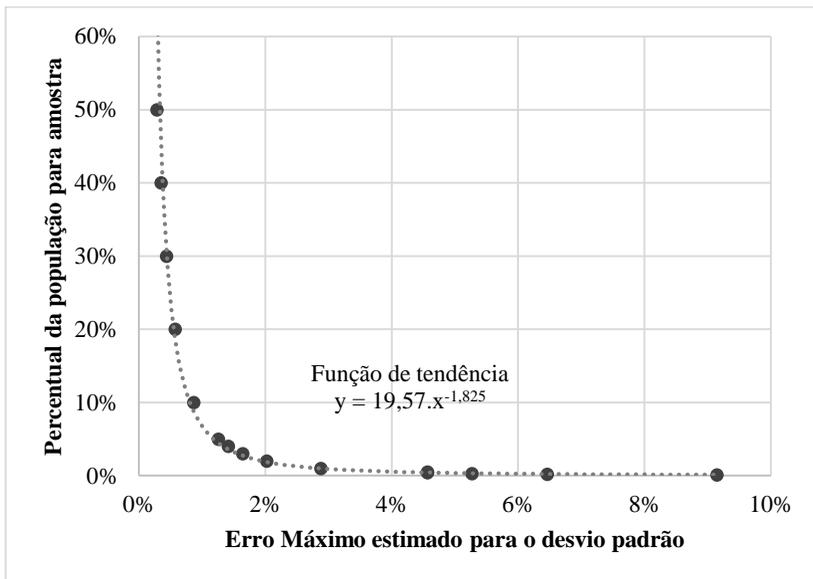


Figura 28. Relação entre o tamanho da amostra e o Erro máximo estimado da amostra

Conforme apresentado na Figura 28, o aumento do tamanho da amostra reduz seu erro máximo estimado em uma função potência. Esta função tende a um erro máximo próximo de zero quanto maior é o tamanho da amostra. Com amostras mais confiáveis, os modelos devem apresentar previsões mais confiáveis, porém o aumento da amostra aumenta a complexidade da modelagem, acarretando em maior tempo de processamento.

5.1.2. NÚMERO DE NÓS NA CAMADA INTERNA

A arquitetura da Rede Neural Artificial adotada foi a de *feed-forward* com uma camada interna. O tamanho da camada interna pode variar em seu número de nós. Não existe embasamento teórico para se determinar o tamanho da camada interna, mas seu dimensionamento está diretamente relacionado ao desempenho da rede. Quanto maior o número de nós, mais pesos para ajustar e mais complexa será a RNA. Quanto maior a complexidade, melhor o desempenho da RNA, entretanto, o maior número de pesos aumenta o tempo de processamento da rede (LAWRENCE; GILES; TSOI, 1996).

Geralmente o tamanho da camada interna é dimensionada em função do número de variáveis preditoras e preditas, dimensão da amostra para treinamento, ruídos da amostra e da complexidade do sistema a representar. Na bibliografia encontram-se algumas “regras de ouro” que são facilmente refutadas (SARLE, 2002), como:

- A camada interna deve ter o tamanho entre o número de nós na camada de entrada e o número de nós da camada de saída;
- O tamanho da camada interna é igual a dois terços da soma de nós da camada de entrada e saída;
- A camada interna nunca deverá ser maior que o dobro de nós da camada de entrada.

Porém, não há maneira melhor de determinar o tamanho da camada interna do que treinando várias redes. É necessário refinar a complexidade do modelo até que os dados estejam bem ajustados em relação aos processos de validação (SARLE, 2002). Para este trabalho foram dimensionados quatro tamanhos de camada interna com número de nós

em intervalos. Foram consideradas as proporções de 0,5, 1, 2 e 3 vezes o número de nós das camadas de entrada e saída. Portanto, como são 24 nós (variáveis preditoras + variáveis preditas), foram consideradas camadas internas com 12, 24, 48 e 72 nós.

5.1.3. AVALIAÇÃO DOS MODELOS

Foram combinados os quinze tamanhos de amostra com os quatro tamanhos de camada interna considerados para este refinamento, resultando em 60 metamodelos de Redes Neurais Artificiais. Para avaliação e comparação destes modelos foram adotados os mesmos três índices de desempenho utilizados no Capítulo 4: tempo de treinamento; coeficiente de determinação de Spearman (R^2); Raiz do erro médio quadrático (RMSE) e Raiz do erro médio quadrático normalizada (NRMSE). O tempo foi registrado durante o treinamento dos modelos e o R^2 , RMSE e NRMSE foram calculados a partir da aplicação dos modelos para predição de toda base de dados, ou seja, para predição dos 1,29 milhões de casos. Para classificação das Redes Neurais Artificiais de melhor desempenho foi calculado o Índice de Desempenho Global (IDG) considerando o tempo de treinamento, o R^2 e a RMSE.

5.1.4. PREDIÇÃO DE PARÂMETROS NÃO VISTOS

Para testar o poder de generalização dos Metamodelos de Redes Neurais Artificiais foram realizados testes para valores de parâmetros não vistos. Mesmo que o desempenho de predição para os 1,29 milhões de casos seja satisfatório, estes possuem os mesmos valores dos parâmetros utilizados para treinamento da rede. Para isso foram simulados novos

casos da zona térmica com valores de parâmetros não utilizados anteriormente. Como as funções de Redes Neurais Artificiais não garantem a extrapolação de valores de parâmetros, foram elaborados dois grupos de valores não vistos: um grupo de valores intermediários aos utilizados (para testar a interpolação); e um grupo de valores que extrapolam os utilizados (para testar a extrapolação).

A Tabela 11 apresenta os valores adotados para os testes de não vistos. A Zona Térmica foi simulada variando cada um dos valores individualmente, e mantendo os outros parâmetros com valores médios já vistos, apresentados como caso base. A Tabela 12 complementa os parâmetros dos casos não vistos com os tipos de coberturas e paredes.

Tabela 11. Valores não vistos intermediários e extrapolados.

Parâmetro	Unidade	Caso base	Valores não vistos intermediários	Valores não vistos extrapolados
Massa Térmica Interna	m ²	26	10; 30; 40	0; 70
Absortância solar da parede externa (aPar)	-	0,4	0,35; 0,74	0,1; 0,9
Absortância solar da cobertura (aCob)	-	0,4	0,35; 0,74	0,1; 0,9
Percentual de janela na fachada (WWR)	%	40	37; 72	90
Fator Solar do vidro (FS)		0,65	0,3; 0,8	0,18
Transmitância térmica do vidro (UVid)	W/m ² .K	5,7	2,0; 4,3	0,3; 1,5
Ângulo vertical de sombreamento (AVS)	°	0	25; 77	85
Ângulo horizontal de sombreamento (AHS)	°	0	25; 77	85
Ângulo de obstrução da vizinhança (AOV)	°	0	25; 77	85
Horário de ocupação (Ocup)	horas	16h	10; 14; 22	4; 6
Densidade de Equip.e Iluminação (DEI)	W/m ²	20	10; 30	0; 60
Densidade de pessoas (DPessoas)	pessoas/m ²	0,4	0,15; 0,55	0,05; 0,85
Orientação Solar (Azimute)	°	180	15; 200	335; 360
Infiltração	ACH	0,6	0,5; 0,75	0,1; 2,0
Pé-direito	m	3,0	2,7; 6	2,4; 8,0
Exposição da Cobertura (Exp_Cob)	-	Externa	Exterior	Exterior
Exposição do Piso (Exp_Piso)	-	Adiabática	Adiabática	Adiabática

Foram definidos tipos de paredes e coberturas comuns para se definir parâmetros vistos e tipos de paredes e coberturas isoladas, leves ou pesadas para se definir parâmetros não vistos. Como parede base foi utilizada a Parede 6, adotada na elaboração da base de dados. A Tabela 12 apresenta os valores de transmitância térmica e capacidade térmica resultantes do tipo de construção adotado para parede ou cobertura.

Tabela 12. Tipos de paredes e coberturas não vistos

Construção	U (W/m².K)	CT (kJ/m².K)
Cobertura Isolada	0,180	8,0
Cobertura Comum	1,166	31,0
Cobertura Leve	5,107	11,0
Cobertura Pesada	2,804	367,0
Parede Isolada	0,190	8,0
Parede Comum	2,734	194,0
Parede Leve	5,724	30,0
Parede Pesada	2,956	427,0
Parede Base	2,781	88,0

A amostra intencional escolhida para os valores não vistos apresentados nas Tabela 11 e Tabela 12 somam 35 valores intermediários e 32 extrapolados. O parâmetro de exposição da cobertura foi fixado como exposição para o exterior e a exposição do piso foi mantida como adiabática. Os casos não vistos foram simulados no programa EnergyPlus com as mesmas configurações dos casos simulados para treinamento dos Metamodelos.

Além dos casos com valores de parâmetros não vistos individualmente, foram simulados dois casos com todos os valores de parâmetros não vistos, um com todos os valores não vistos intermediários e outro com valores não vistos extrapolados. Estes casos foram montados com os valores médios ou maiores de cada grupo de não vistos, destacados em negrito na Tabela 11.

Esses casos foram simulados e preditos pelos metamodelos das arquiteturas de redes neurais artificiais. Foram comparados os valores de correlação e raiz do erro médio quadrático (RMSE) entre o resultado simulado e predito com os valores não vistos.

5.2. RESULTADOS

O trabalho computacional de treinamento das Redes Neurais Artificiais mais simples demoraram poucos segundos, enquanto as redes mais complexas demoraram dias. Entretanto, devido à complexidade, a Rede Neural Artificial com amostra de treinamento de 50% dos resultados simulados e 72 nós na camada interna causou o estouro da pilha de memória RAM, não sendo possível ser modelado com a configuração computacional disponível. Esta era a rede mais complexa a ser modelada e foi descartada das análises. Os outros processamentos foram realizados sem contratempos.

A Tabela 13 apresenta o resumo dos resultados da validação cruzada de todas as Redes Neurais Artificiais treinadas. Para o processo de validação cruzada, cada Rede Neural Artificial foi treinada 64 vezes com diferentes amostragens dentro da amostra separada para treino. Foram calculados os valores de coeficiente de determinação (R^2) e a raiz do erro médio quadrático (RMSE) de cada caso de treinamento e então apresentados os valores mínimos, máximos, médias e medianas para cada arquitetura de Rede Neural Artificial. O R^2 e o RMSE para validação cruzada foram calculados a partir da amostra de 1/8 da amostra separada para teste de cada treinamento. Foi calculada a diferença entre a melhor e a pior Rede Neural Artificial dos 64 treinamentos para cada arquitetura. Essa diferença foi utilizada para validação do modelo, pois quanto menor a diferença entre o melhor e pior modelo, maior é a consistência do tamanho da amostra e número de nós na camada interna da rede. A escala de cores na Tabela 13 representa a grandeza do valor, esta escala inicia com o verde para os valores de melhor desempenho e vermelho para os de pior desempenho, passando por amarelo para os intermediários.

Tabela 13. R² e RMSE dos treinamentos de validação cruzada das Redes Neurais Artificiais

Amostra	Camada interna	RMSE (kWh/m ² .ano)					Spearman R ²				
		Mín.	Mediana	Média	Máx.	Dif. Máx. e Mín.	Mín.	Mediana	Média	Máx.	Dif. entre Máx. e Mín.
0,1%	12	1,100	1,141	1,145	1,193	8%	0,727	0,847	0,843	0,932	22%
	24	1,201	1,244	1,247	1,353	11%	0,464	0,656	0,650	0,781	41%
	48	1,178	1,230	1,232	1,312	10%	0,508	0,668	0,675	0,781	35%
	72	1,149	1,209	1,212	1,292	11%	0,509	0,718	0,703	0,844	40%
0,2%	12	1,076	1,096	1,096	1,124	4%	0,874	0,925	0,923	0,949	8%
	24	1,065	1,087	1,087	1,124	5%	0,873	0,940	0,937	0,965	9%
	48	1,122	1,157	1,157	1,194	6%	0,757	0,827	0,825	0,894	15%
	72	1,158	1,187	1,188	1,233	6%	0,689	0,768	0,764	0,826	17%
0,3%	12	1,071	1,085	1,087	1,116	4%	0,888	0,941	0,939	0,960	7%
	24	1,053	1,064	1,064	1,092	4%	0,938	0,967	0,965	0,975	4%
	48	1,059	1,073	1,073	1,089	3%	0,938	0,957	0,956	0,975	4%
	72	1,103	1,118	1,118	1,148	4%	0,828	0,898	0,896	0,926	11%
0,4%	12	1,072	1,080	1,081	1,096	2%	0,926	0,946	0,946	0,958	3%
	24	1,052	1,059	1,060	1,083	3%	0,944	0,970	0,970	0,977	3%
	48	1,045	1,051	1,051	1,061	2%	0,969	0,977	0,977	0,982	1%
	72	1,056	1,066	1,067	1,098	4%	0,929	0,964	0,963	0,974	5%
0,5%	12	1,071	1,081	1,081	1,096	2%	0,924	0,948	0,947	0,960	4%
	24	1,051	1,056	1,057	1,077	2%	0,950	0,974	0,973	0,981	3%
	48	1,038	1,044	1,044	1,050	1%	0,979	0,984	0,984	0,988	1%
	72	1,042	1,047	1,048	1,054	1%	0,976	0,981	0,981	0,985	1%
1%	12	1,072	1,077	1,078	1,087	1%	0,939	0,953	0,952	0,960	2%
	24	1,046	1,051	1,052	1,060	1%	0,970	0,978	0,978	0,982	1%
	48	1,032	1,035	1,035	1,041	1%	0,986	0,990	0,990	0,991	1%
	72	1,029	1,031	1,031	1,034	1%	0,990	0,992	0,992	0,993	0%
2%	12	1,070	1,077	1,077	1,085	1%	0,943	0,953	0,953	0,962	2%
	24	1,046	1,050	1,051	1,058	1%	0,973	0,980	0,979	0,982	1%
	48	1,031	1,033	1,033	1,037	1%	0,989	0,991	0,991	0,992	0%
	72	1,026	1,028	1,028	1,032	1%	0,992	0,994	0,993	0,995	0%
3%	12	1,069	1,076	1,076	1,089	2%	0,939	0,954	0,954	0,962	2%
	24	1,046	1,050	1,050	1,060	1%	0,972	0,980	0,980	0,982	1%
	48	1,030	1,032	1,032	1,035	0%	0,990	0,991	0,991	0,992	0%
	72	1,025	1,027	1,027	1,028	0%	0,993	0,994	0,994	0,995	0%
4%	12	1,068	1,077	1,077	1,086	2%	0,941	0,953	0,953	0,963	2%
	24	1,046	1,049	1,050	1,059	1%	0,972	0,980	0,980	0,983	1%
	48	1,030	1,032	1,032	1,034	0%	0,990	0,992	0,991	0,993	0%
	72	1,024	1,026	1,026	1,027	0%	0,994	0,994	0,994	0,995	0%
5%	12	1,070	1,079	1,078	1,087	2%	0,941	0,952	0,952	0,961	2%
	24	1,046	1,050	1,050	1,058	1%	0,973	0,980	0,979	0,983	1%
	48	1,029	1,032	1,032	1,037	1%	0,989	0,992	0,992	0,993	0%
	72	1,024	1,026	1,026	1,028	0%	0,994	0,995	0,995	0,995	0%
10%	12	1,070	1,076	1,076	1,092	2%	0,935	0,954	0,953	0,960	3%
	24	1,047	1,049	1,050	1,059	1%	0,972	0,980	0,979	0,982	1%
	48	1,030	1,032	1,032	1,035	1%	0,990	0,992	0,992	0,993	0%
	72	1,024	1,026	1,026	1,028	0%	0,994	0,994	0,994	0,995	0%

Tabela 13. R² e RMSE dos treinamentos de validação cruzada das Redes Neurais Artificiais

Amostra	Camada interna	RMSE (kWh/m ² .ano)					Spearman R ²				
		Mín.	Mediana	Média	Máx.	Dif. Máx. e Mín.	Mín.	Mediana	Média	Máx.	Dif. entre Máx. e Mín.
20%	12	1,068	1,077	1,077	1,090	2%	0,937	0,953	0,952	0,963	3%
	24	1,047	1,050	1,051	1,061	1%	0,970	0,980	0,979	0,982	1%
	48	1,030	1,032	1,032	1,035	0%	0,990	0,992	0,992	0,993	0%
	72	1,025	1,026	1,026	1,027	0%	0,994	0,995	0,995	0,995	0%
30%	12	1,070	1,079	1,078	1,088	2%	0,939	0,951	0,951	0,961	2%
	24	1,046	1,049	1,050	1,059	1%	0,972	0,980	0,980	0,983	1%
	48	1,029	1,032	1,032	1,035	1%	0,990	0,992	0,992	0,993	0%
	72	1,025	1,026	1,026	1,027	0%	0,994	0,994	0,994	0,995	0%
40%	12	1,069	1,076	1,077	1,088	2%	0,939	0,954	0,953	0,963	2%
	24	1,046	1,050	1,050	1,059	1%	0,972	0,980	0,980	0,983	1%
	48	1,030	1,032	1,032	1,035	0%	0,990	0,992	0,992	0,993	0%
	72	1,025	1,026	1,026	1,027	0%	0,994	0,994	0,994	0,995	0%
50%	12	1,071	1,076	1,077	1,084	1%	0,944	0,954	0,953	0,960	2%
	24	1,046	1,049	1,051	1,060	1%	0,971	0,980	0,979	0,983	1%
	48	1,030	1,032	1,032	1,035	0%	0,990	0,992	0,991	0,992	0%

Conforme a Tabela 13, os RMSE apresentam pouca diferença para todas as arquiteturas. Com exceção das arquiteturas com menor número de amostras, de 0,1%. O valor de erro foi utilizado como parâmetro de convergência dos treinamentos, por isso não apresenta grande variação. Por outro lado, o coeficiente de determinação apresenta diferenças maiores, de até 41%. Esta grande diferença significa que, dentre os 64 treinamentos de cada arquitetura de rede, a diferença na amostragem de casos interferiu diretamente no desempenho da regressão. A diferença entre os melhores e piores resultados de treinamento foi utilizada como parâmetro de validação das arquiteturas de Redes Neurais Artificiais, onde diferenças maiores que 5% invalidam a arquitetura por serem consideradas instáveis. Portanto, todos os casos com 0,1 e 0,2% de amostra e os casos de 0,3% de tamanho de amostra com 12 e 72 nós na camada interna não foram consideradas redes válidas. Avaliando os

valores de RMSE e R^2 , os melhores desempenhos são encontrados nas arquiteturas com amostras maiores que 0,5% e com camadas internas de 48 e 72 nós. Nas arquiteturas com amostra menor que 0,3%, os tamanhos da camada interna não melhoraram seu desempenho, confirmando a invalidação destas arquiteturas, pois o aumento da complexidade da rede está diretamente relacionado ao melhor desempenho. O tamanho da amostra é um fator importante na elaboração de qualquer modelo de regressão. Para o uso de Redes Neurais Artificiais com as características adotadas neste trabalho, adotar menos de 3.380 casos, que corresponde a 0,3% do 1,29 milhões de casos simulados, não atinge uma estabilidade satisfatória demonstrada pela validação cruzada.

As Figura 29 e Figura 30 apresentam o tempo de treinamento de cada Rede Neural Artificial. A primeira relaciona o tempo pelo tamanho da amostra e a segunda pelo tamanho da camada interna.

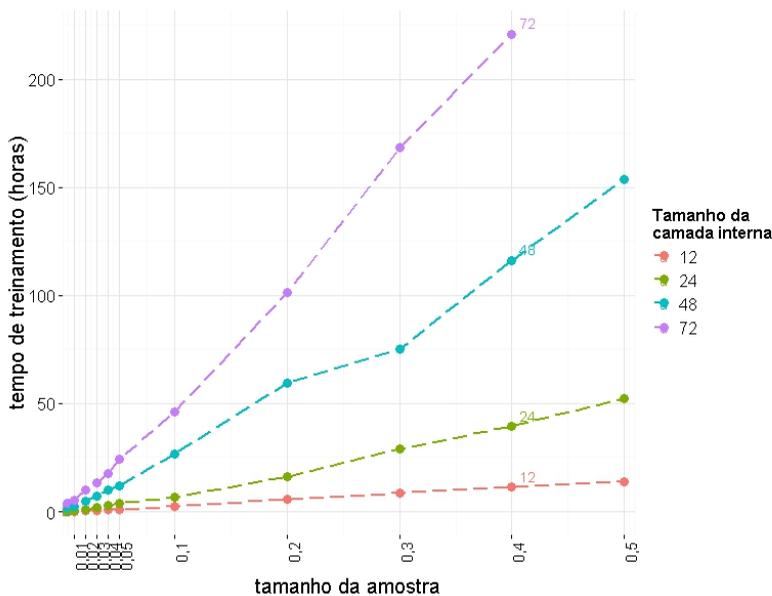


Figura 29. Tempos de treinamento por tamanho de amostra.

Como mostra a Figura 29, o tamanho da amostra tem mais influência no tempo de processamento das redes quando o tamanho da camada interna é maior. A camada interna com 12 nós pouco teve influência no tempo de treinamento em relação ao tamanho da amostra. Enquanto os casos com 72 nós na camada interna teve um acréscimo significativo no tempo de treinamento, levando mais de 220 horas para seu processamento.

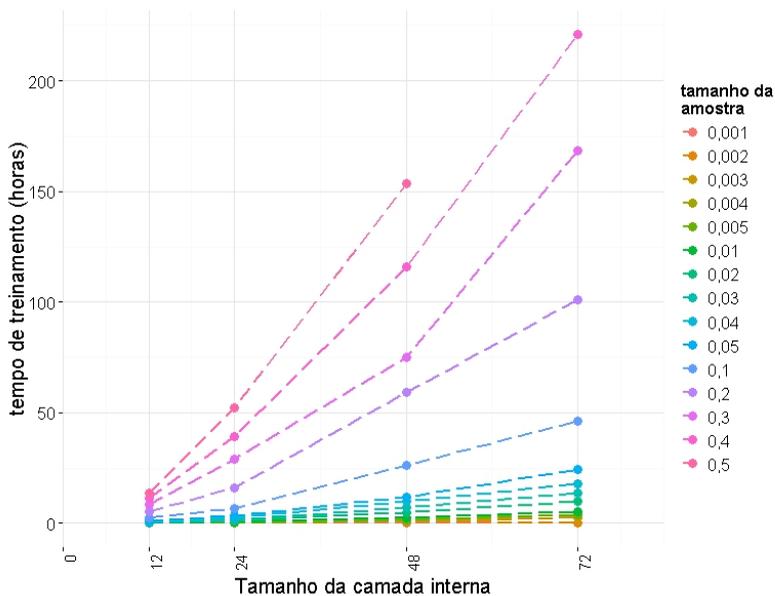


Figura 30. Tempos de treinamento por tamanho da camada interna.

Em relação ao tamanho da camada interna, o gráfico da Figura 30 demonstra como as amostras de tamanho 10 a 50% têm o tempo de processamento acrescido quanto maior a camada interna da rede. Para as redes menores que 10% de amostra, a camada interna tem menos influência para o tempo de treinamento, principalmente para camadas internas menores que 24 nós.

A Figura 31 e Figura 32 apresentam o coeficiente de determinação, R^2 , em relação ao tamanho da amostra e ao tamanho da camada interna, respectivamente.

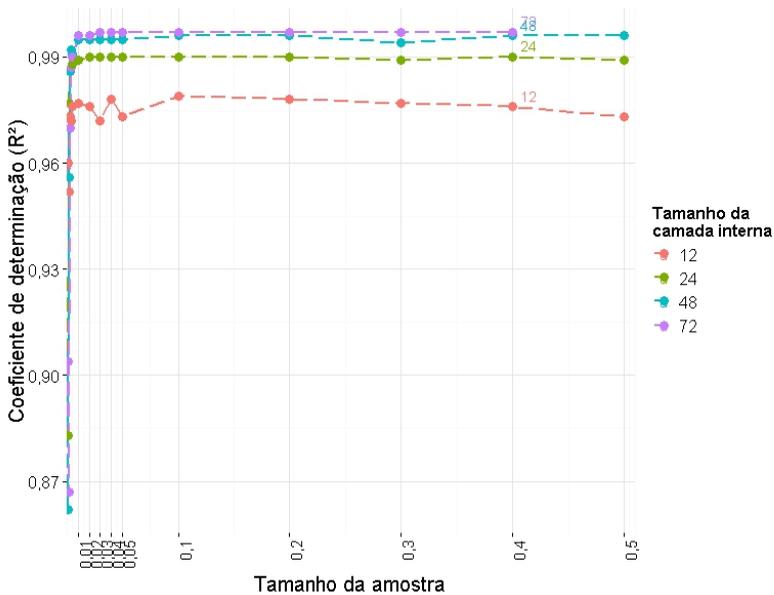


Figura 31. Coeficiente de determinação de Spearman (R^2) por tamanho de amostra.

O tamanho da amostra acima de 1% não apresenta influência para o valor de R^2 de cada tamanho de camada interna, conforme observado na Figura 31, onde as linhas de tamanho de camada interna não variam de valor de R^2 em relação ao tamanho da amostra no eixo X. A não ser pelos casos de 12 nós na camada interna, que apresenta uma variação de R^2 , demonstrando a inconstância dos modelos com este tamanho de camada interna quando com menos de 10% de amostra. Abaixo de 1% de tamanho de amostra, os valores de R^2 decrescem, alcançando valores menores que 0,90.

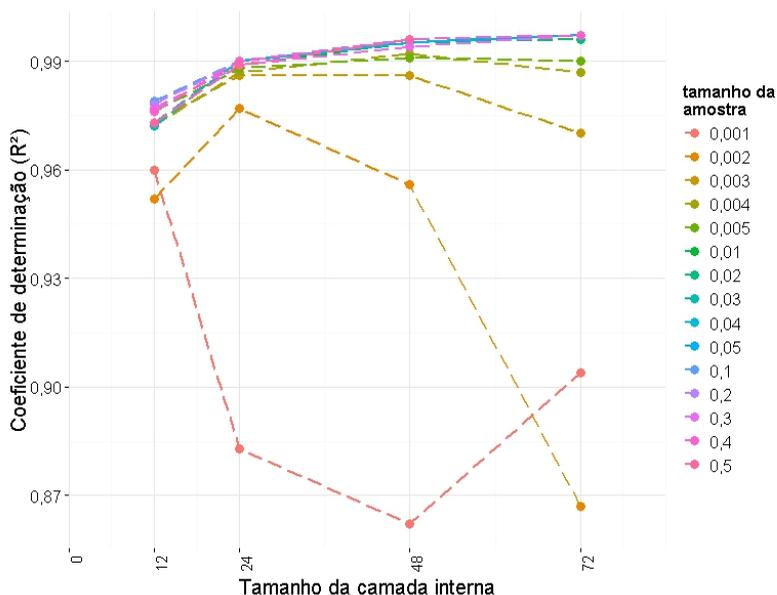


Figura 32. Coeficiente de determinação de Spearman (R^2) por tamanho da camada interna.

A Figura 32 apresenta os valores de R^2 em relação ao tamanho da camada interna por tamanho de amostra. Este gráfico evidencia que para amostras de 0,1 e 0,2%, independentemente do tamanho da camada interna, os valores de R^2 não seguem uma linha crescente, como ocorre com os casos com mais de 0,4%. As arquiteturas com amostras maiores que 0,4% apresentam R^2 muito próximos, aumentando conforme aumenta o tamanho da camada interna. Ou seja, para amostras acima de 0,4%, quanto maior a complexidade da rede, maior o R^2 . Esta observação confirma a invalidação das arquiteturas com menos de 0,4%, conforme o processo de validação cruzada.

A Figura 33 e Figura 34 apresentam os valores da raiz do erro médio quadrático (RMSE) das Redes Neurais Artificiais para predição

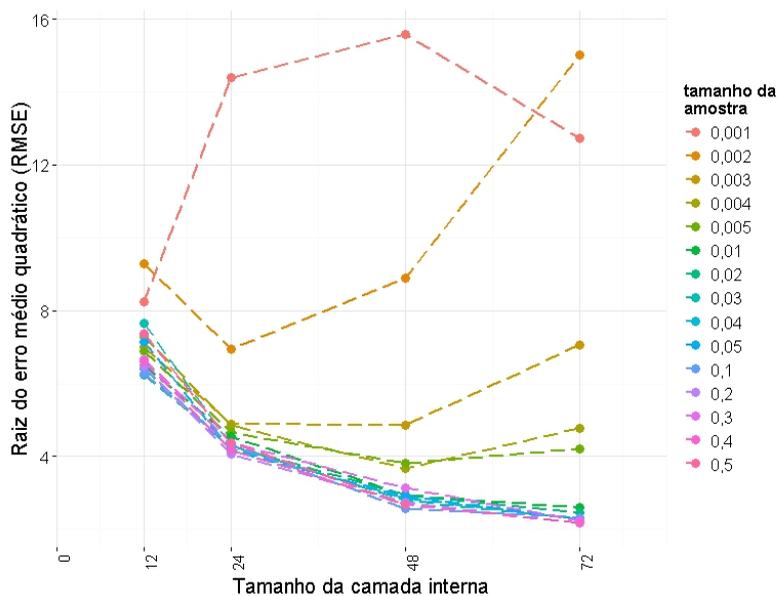


Figura 34. Raiz do erro médio quadrático (RMSE) por tamanho da camada interna.

O gráfico da Figura 34 demonstra como os casos com tamanho de amostra abaixo de 0,5% não apresentam uma proporção linear entre os tamanhos de camada interna. Para as arquiteturas com amostras maiores ou iguais a 1%, quanto maior o tamanho da camada interna, menores são os RMSE.

Analisando os R^2 e os RMSE em relação ao tamanho de amostra e tamanho da camada interna, e descartando os resultados invalidados pelo processo de validação cruzada, com amostras menores que 0,3%, é possível afirmar que o aumento da camada interna tem maior influência no desempenho da Rede Neural do que o aumento da amostra. Ou seja, a partir de um tamanho de amostra suficiente é melhor aumentar o número de neurônios na camada interna do que aumentar o número de casos para treinamento. Entretanto, o tempo demandado para treinamento das Redes

Neurais Artificiais tem influência do tamanho da amostra e do tamanho da camada interna, principalmente com mais de 24 nós na camada interna. Além dos desempenhos de precisão como o R^2 e o RMSE, o tempo de treinamento é uma importante variável pois está relacionada aos recursos para desenvolvimento de um Metamodelo. Portanto, estes fatores de desempenho devem ser ponderados para se determinar qual a arquitetura mais adequada para o desenvolvimento do Metamodelo.

A Tabela 14 apresenta o desempenho de cada arquitetura de rede neural artificial avaliada pelo Índice de Desempenho Global (IDG). Os valores de RMSE e R^2 foram calculados a partir dos resultados da aplicação dos Metamodelos para predição dos 1,29 milhões de casos. O tempo de treino é o tempo computacional demandado para elaboração de cada rede. O Índice de Desempenho Global (IDG) foi calculado a partir do tempo de treinamento, R^2 e RMSE. Os valores de tempo e RMSE tiveram seus sinais invertidos para orientar o maior desempenho para os maiores valores. O valor de IDG foi normalizado para comparação dos valores entre 0 e 1, gerando o IDG normal. Os valores de IDG normal maiores que 0,90, que representam as arquiteturas com maior desempenho, estão destacados em negrito.

Tabela 14. Índices de desempenho das Redes Neurais Artificiais

Amostra	Camada Interna	R ²	RMSE (kWh/m ² .ano)	Tempo de Treino (horas)	IDG	IDG normal
0,1%	12	0,960	-8,240	-0,01	0,758	0,657
	24	0,883	-14,405	-0,04	0,414	0,125
	48	0,862	-15,566	-0,04	0,333	0,000
	72	0,904	-12,734	-0,08	0,507	0,270
0,2%	12	0,952	-9,275	-0,03	0,712	0,587
	24	0,977	-6,932	-0,12	0,832	0,772
	48	0,956	-8,896	-0,68	0,730	0,615
	72	0,867	-15,015	-0,17	0,359	0,040
0,3%	12	0,973	-6,995	-0,04	0,821	0,755
	24	0,986	-4,895	-0,17	0,905	0,885
	48	0,986	-4,849	-0,88	0,905	0,885
	72	0,970	-7,057	-2,58	0,808	0,735
0,4%	12	0,972	-7,278	-0,06	0,811	0,740
	24	0,987	-4,865	-0,23	0,908	0,890
	48	0,992	-3,660	-1,20	0,949	0,953
	72	0,987	-4,766	-4,01	0,905	0,885
0,5%	12	0,976	-6,894	-0,08	0,830	0,770
	24	0,988	-4,656	-0,25	0,915	0,902
	48	0,991	-3,804	-1,43	0,942	0,943
	72	0,990	-4,217	-3,78	0,926	0,918
1,0%	12	0,977	-6,508	-0,18	0,842	0,788
	24	0,989	-4,541	-0,56	0,920	0,909
	48	0,995	-2,934	-2,52	0,972	0,989
	72	0,996	-2,594	-5,25	0,979	1,000
2,0%	12	0,976	-6,241	-0,37	0,846	0,794
	24	0,990	-4,302	-1,17	0,928	0,921
	48	0,995	-2,932	-4,68	0,969	0,984
	72	0,996	-2,451	-10,06	0,975	0,994
3,0%	12	0,972	-7,651	-0,64	0,801	0,724
	24	0,990	-4,291	-1,80	0,927	0,920
	48	0,995	-2,805	-7,08	0,968	0,984
	72	0,997	-2,295	-13,50	0,976	0,996
4,0%	12	0,978	-6,398	-0,92	0,846	0,795
	24	0,990	-4,197	-2,69	0,928	0,921
	48	0,995	-2,846	-10,16	0,963	0,975
	72	0,997	-2,302	-17,70	0,970	0,986
5,0%	12	0,973	-7,153	-1,08	0,815	0,746
	24	0,990	-4,153	-3,69	0,928	0,921
	48	0,995	-2,899	-12,02	0,959	0,969
	72	0,997	-2,285	-24,43	0,960	0,971
10,0%	12	0,979	-6,255	-2,39	0,850	0,801
	24	0,990	-4,318	-6,90	0,919	0,907
	48	0,996	-2,582	-26,43	0,947	0,951
	72	0,997	-2,316	-46,12	0,927	0,919
20,0%	12	0,978	-6,466	-5,54	0,838	0,781
	24	0,990	-4,056	-16,00	0,912	0,896
	48	0,996	-2,738	-59,50	0,893	0,868
	72	0,997	-2,183	-101,16	0,847	0,795

Tabela 14. Índices de desempenho das Redes Neurais Artificiais

Amostra	Camada Interna	R ²	RMSE (kWh/m ² .ano)	Tempo de Treino (horas)	IDG	IDG normal
30,0%	12	0,977	-6,668	-8,48	0,826	0,763
	24	0,989	-4,367	-29,13	0,882	0,849
	48	0,994	-3,131	-75,05	0,855	0,808
	72	0,997	-2,229	-168,62	0,744	0,636
40,0%	12	0,976	-6,614	-11,27	0,820	0,755
	24	0,990	-4,190	-39,40	0,873	0,836
	48	0,996	-2,676	-116,13	0,810	0,738
	72	0,997	-2,165	-220,82	0,667	0,516
50,0%	12	0,973	-7,361	-13,70	0,791	0,709
	24	0,989	-4,391	-52,24	0,846	0,794
	48	0,996	-2,687	-153,70	0,753	0,649

Os resultados de R² e RMSE das Redes Neurais Artificiais para a predição de toda população de 1,29 milhões de casos representa o desempenho dos Metamodelos para uma amostra significativa do universo de possibilidades construtivas e de ocupação de uma zona térmica de edificação comercial. A Tabela 14 destaca as melhores arquiteturas de rede a partir do Índice Global de Desempenho (IDG), que pondera o tempo de treinamento, o R² e o RMSE. As arquiteturas com amostras de 0,5% a 10% e camadas internas de 24, 48 e 72 nós apresentaram os melhores desempenhos. Além destes, com uma amostra menor, a arquitetura de amostra 0,4% e 48 nós na camada interna, também apresentou bom desempenho, com IDG normal de 0,953. As arquiteturas com amostras maiores, com 20 a 50%, e camadas internas maiores de 12 nós, apresentam altos valores de R² e baixos RMSE, entretanto não foram bem avaliadas pelo IDG por levarem de 16 a 220 horas para treinamento. A melhor arquitetura avaliada pelo IDG foi a com amostra de 1%, que representa 12.933 casos, com 72 nós na camada interna. Esta arquitetura levou 5,25 horas para ser treinada e resultou em um R² de 0,996 e 2,594kWh/m².ano de RMSE.

Para visualizar as tendências dos índices de desempenho adotados de tempo de treinamento, R^2 e RMSE, foram elaborados gráficos com superfícies de regressão entre os índices e o tamanho da camada interna e tamanho da amostra. Para isso, foram descartadas as arquiteturas invalidadas pelo processo de validação cruzada, com tamanho de amostra menor que 0,3%. A Figura 35 apresenta a superfície de regressão para os tempos de treinamento em relação ao tamanho da amostra e da camada interna. A Figura 36 apresenta a superfície de regressão para o valor de R^2 e a Figura 36 para o RMSE.

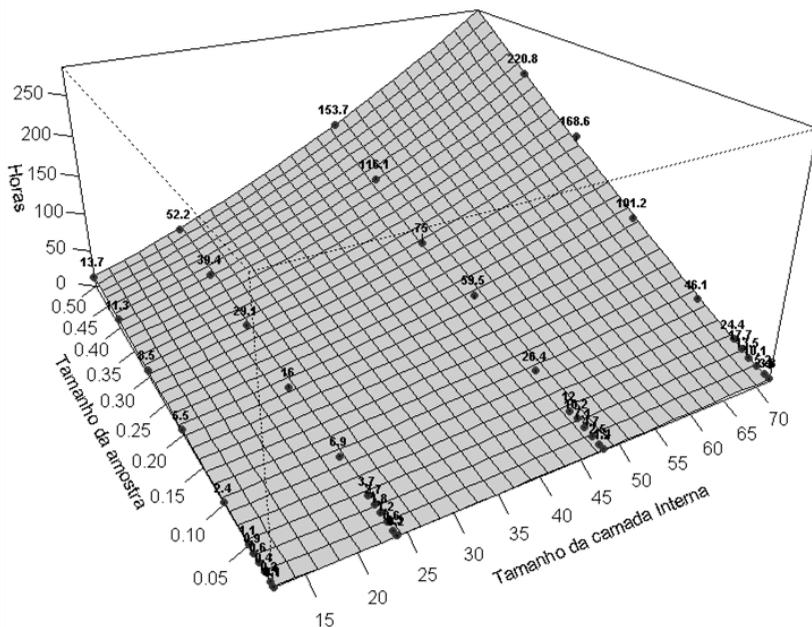


Figura 35. Tempo de treinamento por tamanho de amostra e camada interna.

Com a curva de regressão da Figura 35 é possível estimar que, caso tivesse sido processada, a Rede Neural Artificial com 50% de amostra e 74 nós na camada interna levaria mais de 250 horas para treinamento.

Tanto para o tamanho da amostra, quanto para camada interna, quanto menores os valores, menos tempo de treinamento é necessário. Entretanto, o aumento da complexidade da arquitetura da rede aumenta diretamente o tempo de treinamento.

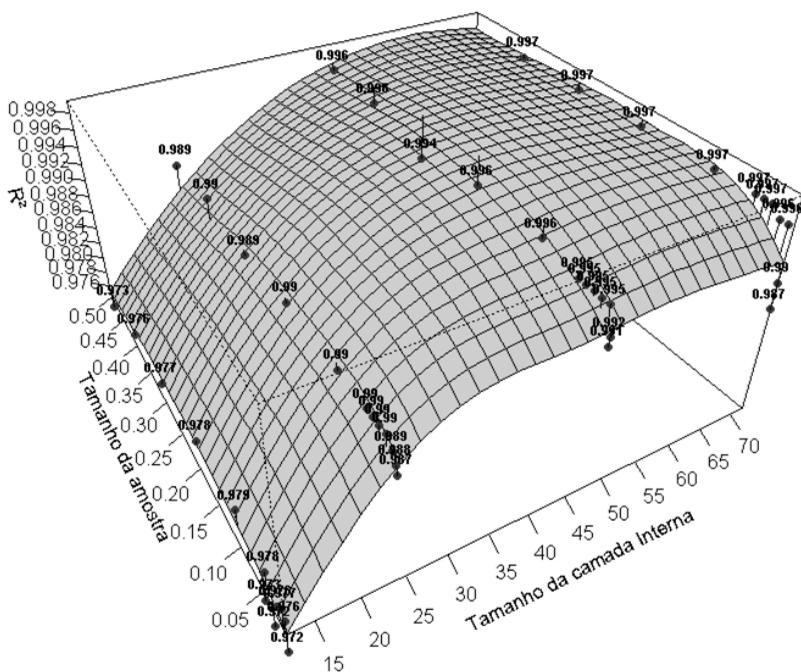


Figura 36. Coeficiente de determinação por tamanho de amostra e camada interna.

Na Figura 36 é possível observar a tendência dos valores de R^2 se estabilizarem próximos a 0,99, onde a superfície forma um platô para os resultados de tamanho de amostra entre 0,1% e 0,5% e tamanho da camada interna entre 48 e 72 nós.

5.2.1. PREDIÇÃO DE CASOS NÃO VISTOS

Os casos com parâmetros não vistos foram simulados e seus resultados foram comparados com os resultados de previsões de todas as Redes Neurais Artificiais treinadas com diferentes arquiteturas. As Figura 38 e Figura 39 apresentam a correlação dos resultados de carga anual de refrigeração das simulações versus os resultados das previsões da Rede Neural Artificial com arquitetura com 1% da amostra e 72 nós na camada interna.

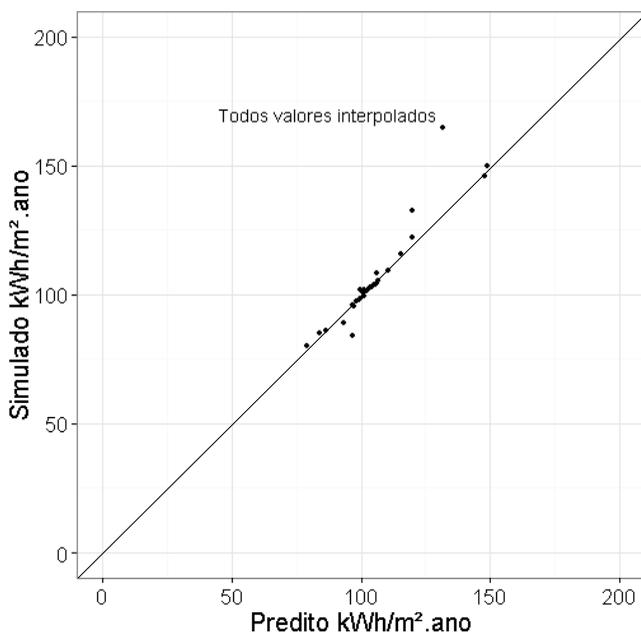


Figura 38. Dispersão dos valores Simulados vs. Preditos dos casos interpolados

A Figura 38 apresenta a dispersão entre os resultados simulados e preditos de carga anual de refrigeração dos casos com valores interpolados. Estes casos apresentaram um coeficiente de determinação,

R^2 , de 0,938 e raiz do erro médio quadrático, RMSE, de 6,959 kWh/m².ano. Este desempenho é pouco menor que o desempenho da Rede Neural Artificial para predição de toda base, que apresentou R^2 de 0,992 e RMSE de 3,660 kWh/m².ano. Do mesmo modo, destaca-se no gráfico o maior erro com o caso com todos os valores dos parâmetros interpolados. Estas observações permitem a conclusão que o desempenho da Rede Neural Artificial da arquitetura com melhor desempenho para os casos vistos tem seu desempenho reduzido para casos não vistos. A seguir, a Figura 39 apresenta o mesmo tipo de análise, porém com os casos com valores extrapolados.

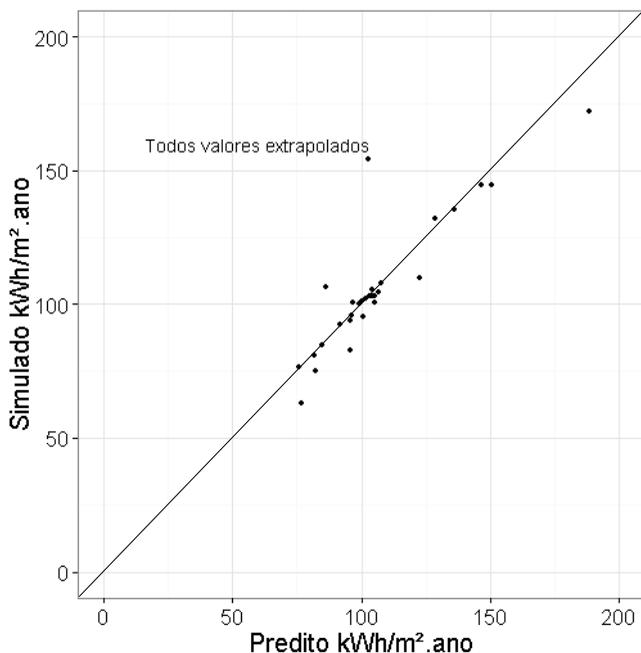


Figura 39. Dispersão dos valores Simulados vs. Preditos dos casos extrapolados.

A avaliação dos casos com valores extrapolados, apresentados no gráfico de dispersão da Figura 39, apresenta um coeficiente de determinação, R^2 , de 0,875 e raiz do erro médio quadrático de 11,597 kWh/m².ano. Estes índices estão aquém dos desempenhados pela Rede Neural Artificial para predição dos casos não vistos com valores interpolados e ainda menores que o desempenho para os valores de toda base. Estes resultados confirmam a premissa de que as Redes Neurais Artificiais não representam bem resultados para valores extrapolados. No gráfico, ainda é destacado o caso com maior erro, onde está identificado como o caso com todos os valores extrapolados.

Para comparar o poder de interpolação e extrapolação das arquiteturas das Redes Neurais Artificiais, os resultados simulados e

preditos foram comparados em termos de correlação pelo coeficiente de determinação, R^2 , e pela raiz do erro médio quadrático, RMSE, para todas as arquiteturas treinadas. Estes índices foram calculados para os grupos de casos com valores interpolados e extrapolados separadamente, os resultados estão apresentados na Tabela 15. Na tabela, os índices foram destacados com uma escala de cores para leitura dos melhores e piores desempenhos. A escala varia do verde para os melhores desempenhos até o vermelho para os piores desempenhos, passando pelo amarelo para desempenhos intermediários.

Tabela 15. Resultados de R^2 e RMSE dos valores simulados vs. preditos para os casos Interpolados e Extrapolados.

Amostra	Camada interna	Interpolados		Extrapolados	
		R^2	RMSE	R^2	RMSE
0,10%	12	0,859	-9,506	0,792	-15,233
	24	0,792	-15,061	0,626	-24,065
	48	0,665	-13,887	0,655	-19,657
	72	0,578	-23,080	0,688	-25,358
0,20%	12	0,720	-13,263	0,829	-18,449
	24	0,906	-8,216	0,831	-14,582
	48	0,889	-8,438	0,859	-14,293
	72	0,835	-11,705	0,585	-30,563
0,30%	12	0,736	-14,605	0,831	-20,244
	24	0,903	-9,440	0,834	-14,415
	48	0,966	-9,621	0,831	-15,874
	72	0,875	-15,245	0,889	-14,757
0,40%	12	0,829	-10,006	0,815	-14,104
	24	0,890	-8,652	0,831	-15,705
	48	0,974	-9,219	0,898	-13,312
	72	0,944	-6,932	0,791	-17,146
0,50%	12	0,830	-10,102	0,789	-15,400
	24	0,911	-8,454	0,896	-11,707
	48	0,912	-7,558	0,872	-11,649
	72	0,940	-6,241	0,887	-11,413
1,00%	12	0,767	-11,357	0,766	-17,244
	24	0,874	-9,063	0,850	-13,267
	48	0,957	-6,198	0,878	-11,164
	72	0,939	-6,528	0,881	-11,260

Tabela 15. Resultados de R^2 e RMSE dos valores simulados vs. preditos para os casos Interpolados e Extrapolados.

Amostra	Camada interna	Interpolados		Extrapolados	
		R^2	RMSE	R^2	RMSE
2,00%	12	0,889	-8,312	0,784	-16,169
	24	0,884	-10,466	0,869	-13,608
	48	0,925	-7,713	0,867	-12,854
	72	0,937	-7,065	0,851	-12,785
3,00%	12	0,668	-18,322	0,701	-27,242
	24	0,938	-10,410	0,849	-16,001
	48	0,921	-7,395	0,876	-12,001
	72	0,947	-6,869	0,863	-12,213
4,00%	12	0,751	-12,037	0,857	-13,845
	24	0,880	-9,570	0,875	-12,652
	48	0,936	-7,787	0,874	-12,017
	72	0,938	-6,959	0,875	-11,597
5,00%	12	0,684	-19,780	0,763	-23,276
	24	0,902	-10,786	0,863	-14,322
	48	0,941	-7,594	0,852	-13,343
	72	0,949	-6,641	0,885	-11,141
10,00%	12	0,784	-12,070	0,785	-19,745
	24	0,884	-8,632	0,865	-12,622
	48	0,955	-7,667	0,879	-11,885
	72	0,943	-6,710	0,898	-10,642
20,00%	12	0,718	-15,725	0,821	-20,658
	24	0,938	-12,249	0,850	-16,195
	48	0,941	-6,966	0,847	-13,143
	72	0,933	-7,334	0,847	-12,713
30,00%	12	0,761	-13,099	0,795	-17,824
	24	0,936	-7,756	0,835	-13,984
	48	0,906	-7,710	0,865	-13,463
	72	0,933	-6,957	0,892	-10,685
40,00%	12	0,804	-11,497	0,815	-16,166
	24	0,879	-8,534	0,892	-11,237
	48	0,929	-8,423	0,858	-13,291
	72	0,249	-24,843	0,865	-13,374
50,00%	12	0,703	-18,910	0,666	-25,127
	24	0,892	-8,138	0,855	-13,086
	48	0,940	-7,922	0,909	-10,509

Nos resultados apresentados na Tabela 15, as médias dos resultados de R^2 para os casos com valores interpolados são maiores que para os casos com valores extrapolados. O maior valor de R^2 foi de 0,974, dado para os casos de valores interpolados para a arquitetura de amostra

com 0,4% e 48 nós na camada interna. O maior valor de R^2 encontrado para os casos com valores extrapolados é de 0,909, para a arquitetura com 50% da amostra e 48 nós na camada interna. Estes resultados demonstram o poder de generalização das Redes Neurais Artificiais. Por outro lado, também se destacam as arquiteturas com pouco poder de generalização, como é caso da Rede Neural Artificial treinada com amostra de 40% e 72 nós na camada interna com R^2 de 0,249 e RMSE de 24,843 kWh/m².ano. Como desempenho para representar os resultados de toda base, esta mesma arquitetura apresentou R^2 de 0,997 e RMSE de 2,165 kWh/m².ano, demonstrando como uma Rede Neural Artificial pode representar bem casos com valores conhecidos mas perde seu desempenho de predição para casos com valores não vistos.

Considerando a importância do desempenho de um Metamodelo para quaisquer valores dentro de limites de contorno de aplicação, foram calculados novos Índices de Desempenho Global (IDG) para as arquiteturas de Redes Neurais Artificiais treinadas. Como limites de aplicação das Redes Neurais Artificiais foram considerados os valores extremos de cada parâmetro da base de dados, ou seja, serão considerados apenas casos com valores intermediários. O IDG foi calculado considerando o tempo de treinamento e os valores de desempenho de R^2 e RMSE das arquiteturas para toda base e para os casos não vistos. A Tabela 16 apresenta os resultados dos Índices de Desempenho Global apenas para as arquiteturas válidas pelo processo de validação cruzada, com amostras maiores que 0,3%. Para simplificar, estão representados apenas os valores normalizados de cada parâmetro de desempenho e não os valores reais. A média dos valores normalizados é o Índice de Desempenho Global utilizado para comparação das arquiteturas.

Tabela 16. Índices de Desempenho Global com desempenho para casos interpolados.

Amostra	Camada interna	Tempo de Treino	R ² para toda base	RMSE para toda base	R ² para valores interpolados	RMSE para valores interpolados	IDG
0,4%	12	1,00	0,81	0,62	0,80	0,80	0,806
	24	1,00	0,93	0,80	0,88	0,87	0,895
	48	0,99	0,96	0,89	1,00	0,84	0,937
	72	0,98	0,93	0,81	0,96	0,96	0,926
0,5%	12	1,00	0,84	0,65	0,80	0,79	0,817
	24	1,00	0,93	0,81	0,91	0,88	0,908
	48	0,99	0,96	0,88	0,91	0,93	0,934
	72	0,98	0,95	0,85	0,95	1,00	0,946
1,0%	12	1,00	0,85	0,68	0,71	0,72	0,793
	24	1,00	0,94	0,82	0,86	0,85	0,894
	48	0,99	0,99	0,94	0,98	1,00	0,979
	72	0,98	0,99	0,97	0,95	0,98	0,974
2,0%	12	1,00	0,84	0,70	0,88	0,89	0,861
	24	0,99	0,95	0,84	0,88	0,77	0,886
	48	0,98	0,99	0,94	0,93	0,92	0,952
	72	0,95	0,99	0,98	0,95	0,95	0,966
3,0%	12	1,00	0,81	0,59	0,58	0,35	0,666
	24	0,99	0,95	0,84	0,95	0,77	0,901
	48	0,97	0,99	0,95	0,93	0,94	0,954
	72	0,94	1,00	0,99	0,96	0,96	0,971
4,0%	12	1,00	0,86	0,68	0,69	0,69	0,784
	24	0,99	0,95	0,85	0,87	0,82	0,895
	48	0,95	0,99	0,95	0,95	0,91	0,950
	72	0,92	1,00	0,99	0,95	0,96	0,964
5,0%	12	1,00	0,82	0,63	0,60	0,27	0,663
	24	0,98	0,95	0,85	0,90	0,75	0,887
	48	0,95	0,99	0,95	0,95	0,93	0,951
	72	0,89	1,00	0,99	0,97	0,98	0,964
10,0%	12	0,99	0,87	0,69	0,74	0,69	0,795
	24	0,97	0,95	0,84	0,88	0,87	0,900
	48	0,88	0,99	0,97	0,97	0,92	0,947
	72	0,79	1,00	0,99	0,96	0,97	0,942
20,0%	12	0,97	0,86	0,68	0,65	0,49	0,730
	24	0,93	0,95	0,86	0,95	0,68	0,872
	48	0,73	0,99	0,96	0,95	0,96	0,919
	72	0,54	1,00	1,00	0,94	0,94	0,885
30,0%	12	0,96	0,85	0,66	0,71	0,63	0,763
	24	0,87	0,94	0,84	0,95	0,92	0,902
	48	0,66	0,98	0,93	0,91	0,92	0,878
	72	0,24	1,00	1,00	0,94	0,96	0,827
40,0%	12	0,95	0,84	0,67	0,77	0,72	0,789
	24	0,82	0,95	0,85	0,87	0,87	0,873
	48	0,47	0,99	0,96	0,94	0,88	0,849
	72	0,00	1,00	1,00	0,00	0,00	0,400

Tabela 16. Índices de Desempenho Global com desempenho para casos interpolados.

Amostra	Camada interna	Tempo de Treino	R ² para toda base	RMSE para toda base	R ² para valores interpolados	RMSE para valores interpolados	IDG
50,0%	12	0,94	0,82	0,61	0,63	0,32	0,663
	24	0,76	0,94	0,83	0,89	0,90	0,864
	48	0,30	0,99	0,96	0,95	0,91	0,824

Com a ponderação dos desempenhos para os casos vistos e não vistos e considerando o tempo de treinamento, os Índices de Desempenho Global apresentados na Tabela 16 foram utilizados para comparar as arquiteturas quanto a seu poder de predição e tempo de elaboração. Com até 5% de amostra, as arquiteturas com melhor desempenho foram as com 48 e 72 nós na camada interna, e da arquitetura com 3% de amostra e 24 nós na camada interna. As arquiteturas com 40 e 50% da amostra apresentam menores desempenhos principalmente pela alta demanda de tempo para treinamento das Redes Neurais Artificiais com esta complexidade. As melhores arquiteturas de Rede Neural Artificial, para os termos aqui analisados, foram as com amostra de 1% e 48 ou 72 nós na camada interna. Dentre estes, o melhor desempenho apresentado pela arquitetura com 72 nós na camada interna coincide com a arquitetura de melhor desempenho para predição de toda base, apresentada na Tabela 14. Desta forma, a Rede Neural Artificial de arquitetura com amostra de 1% e com 72 nós na camada interna apresenta maior desempenho global para predição de todos os valores da base, com valores vistos, e para os casos com valores não vistos interpolados.

5.2.2. ANÁLISE DOS ERROS DA ARQUITETURA DA RNA SELECIONADA

Como resultado da análise dos principais parâmetros que definem a complexidade da arquitetura da Rede Neural Artificial, foi definido que a arquitetura treinada com 1% de amostra da base de simulações e com 72 nós na camada interna foi a combinação com melhor desempenho global. A amostra de 1% da base de simulações equivale a 12.933 casos, com um erro amostral de 0,835kWh/m².ano, ou 2,88%. O Metamodelo não pode superar esse erro, pois este é inerente à amostra utilizada para elaboração do modelo de predição. O Metamodelo produzido com a arquitetura selecionada levou 5,25 horas de processamento computacional, alcançando um coeficiente de determinação de 0,996 e raiz do erro médio quadrático de 2,594kWh/m².ano para predição de todos os casos da base de simulação. Foram analisados os desvios que causam os erros da predição através da dispersão dos resultados simulados pelo preditos e com o histograma de frequência dos erros, com bandas de desvios de 5kWh/m².ano, apresentados na Figura 40.

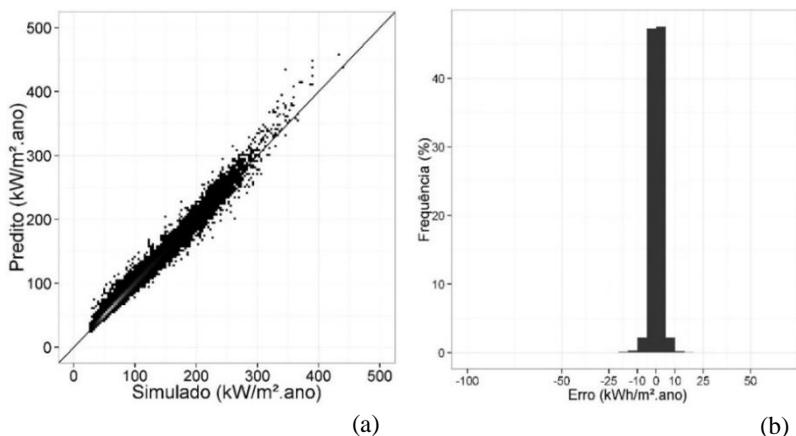


Figura 40. Dispersão dos resultados preditos vs. simulados e histograma dos erros.

O gráfico de dispersão da Figura 40(a) destaca a concentração de pontos com uma cor mais clara. Os casos desta maior concentração apresentam forte correspondência entre os valores preditos e simulados, com uma densidade maior de casos com resultados entre 50 e 150 kWh/m².ano. Existem pontos com desvios positivos e negativos, distribuindo-se de forma aproximadamente simétrica à linha de correspondência perfeita, representada pela diagonal do gráfico. O histograma dos erros, Figura 40(b), quantifica a concentração dos erros próximos a zero, onde 94,5% dos casos apresentam erros absolutos entre -5 e 5 kWh/m².ano. Foram identificados 32.124 casos com erros maiores que 5 kWh/m².ano e 34.728 casos com erros menores que 5 kWh/m².ano. Os demais 1.226.398 casos apresentam erros entre -5 e 5 kWh/m².ano. Ou seja, 66.852 casos apresentam desvios maiores que 5 kWh/m².ano, sendo positivos ou negativos, o que representa apenas 5,5% dos casos da base de dados simulados.

A Figura 41 apresenta a frequência de ocorrência dos erros relativos dos casos, o percentual entre o desvio da predição e o resultado

da simulação. Apesar de semelhante, este gráfico representa valores diferentes do histograma da Figura 40. A análise do erro relativo permite visualizar a ocorrência de erros em relação a seu valor real, simulado.

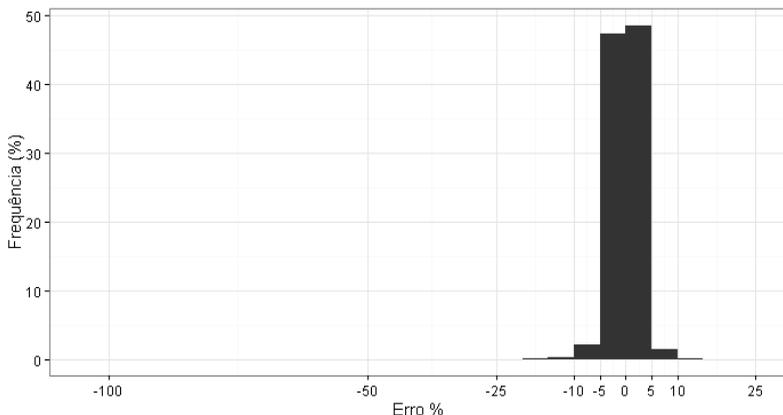


Figura 41. Histograma dos erros relativos

A Figura 41 apresenta o histograma dos erros relativos, que assim como os erros absolutos, concentram-se próximos de zero. O maior erro positivo encontrado foi de 19% e o maior erro negativo foi de -95%. Entretanto, foram identificados apenas 555 casos com desvios percentuais maiores que 20%. Estes casos representam apenas 0,04% da base de simulações. Os erros absolutos positivos são maiores para os casos com cargas anuais de refrigeração maiores. Isto é ilustrado pelo fato do maior erro absoluto positivo ser de 60,95kWh/m².ano e o maior erro relativo positivo ser de 19%. Ou seja, o erro positivo de até 19% é de valores de carga alta, com erro absoluto de 60,95kWh/m².ano. Assim, não se pode afirmar que a ocorrência de erros é diferenciada pela grandeza da carga anual de refrigeração, já que a análise é feita para o percentual de erro.

Foram encontrados 9.974 casos com desvios maiores que 10%, tanto positivos quanto negativos. Estes casos representam 0,8% da base

de simulação. A Figura 42 apresenta a dispersão dos resultados preditos versus resultados simulados, destacando três faixas de erros relativos: erro negativo menor que -10%; erro positivo maior que 10%; e erros entre -10 e 10%.

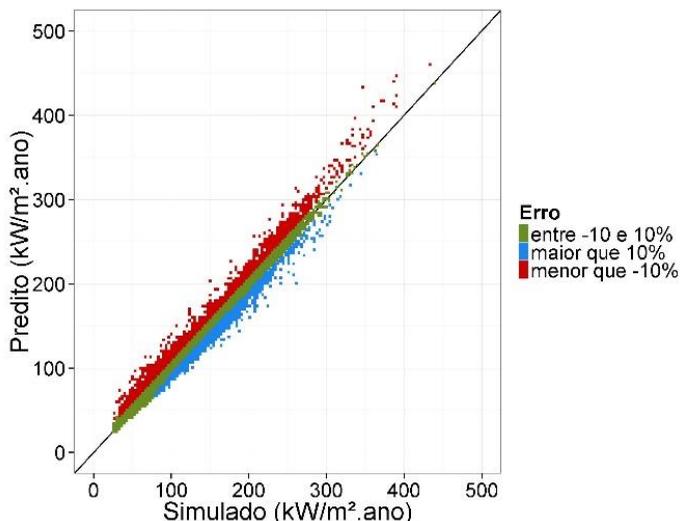
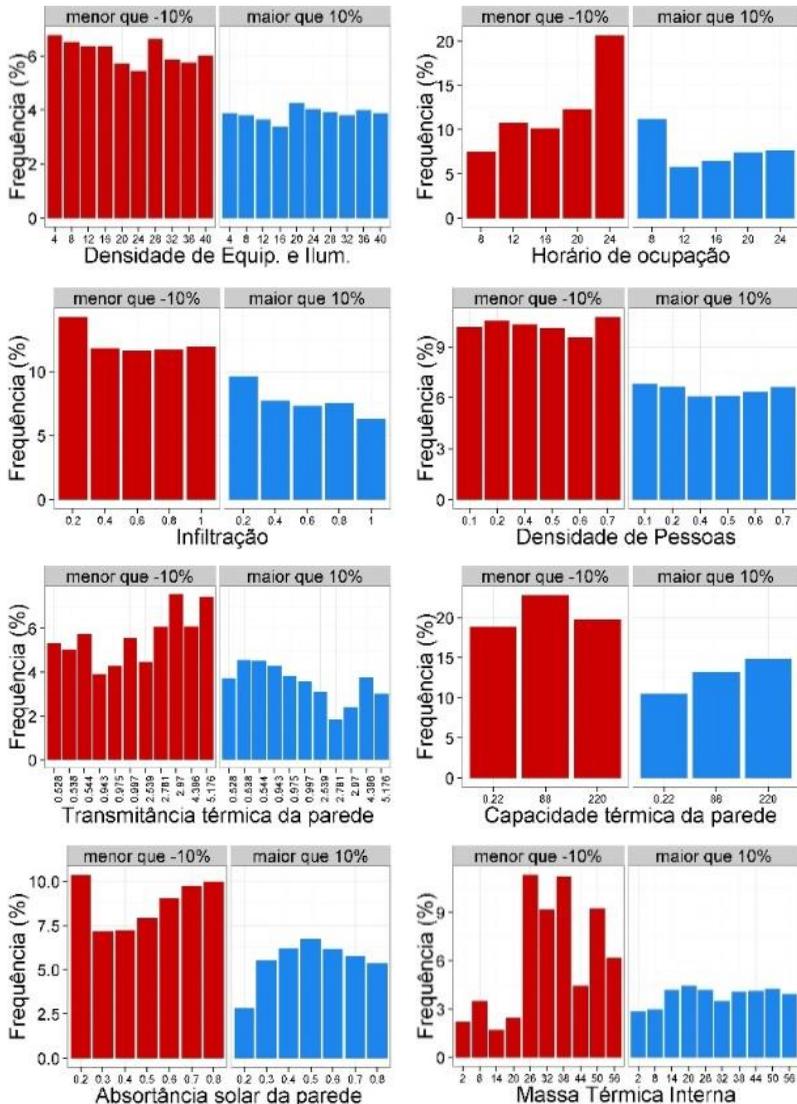


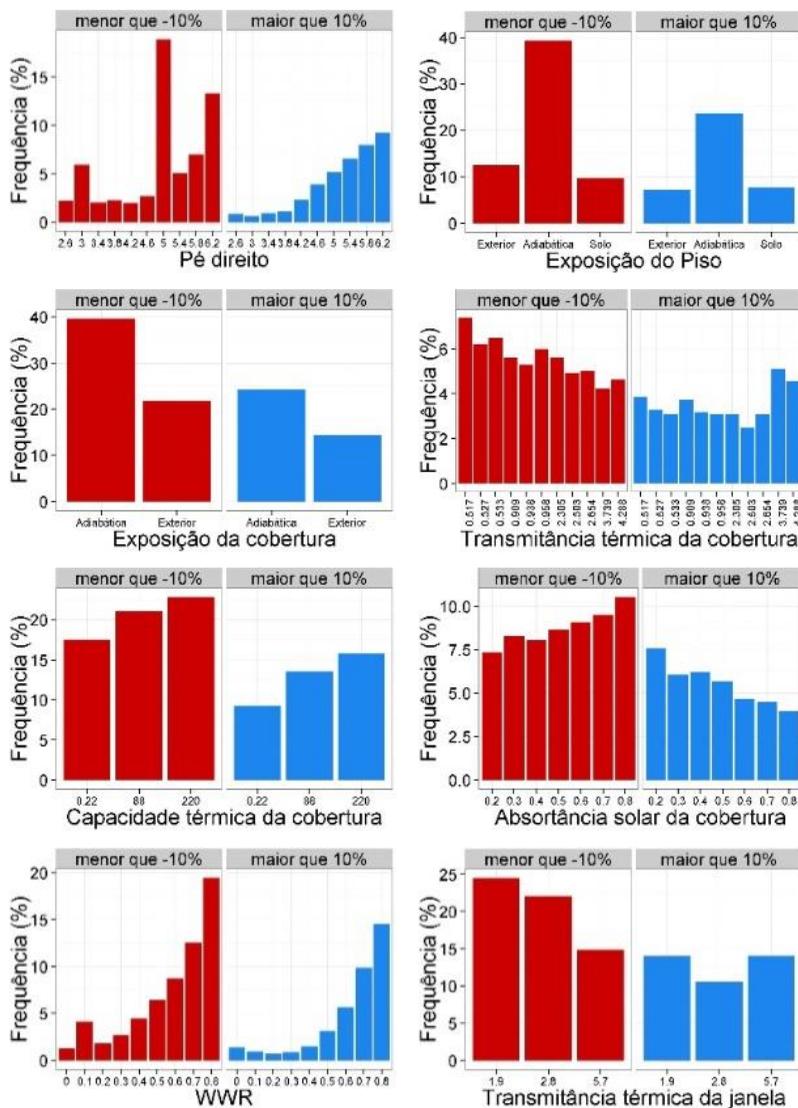
Figura 42. Dispersão entre resultados simulados e preditos por faixa de erro relativo.

Os casos com erros percentuais de no máximo 10% representam 99,2% de todos os casos. Os casos com erros maiores que 10% se afastam da diagonal de correspondência perfeita. A dispersão demonstra a tendência de mais erros negativos que positivos, principalmente para casos com carga anual de refrigeração menor que 100kWh/m².ano e maior que 300kWh/m².ano, onde destaca-se mais pontos com desvio maior que 10% acima da reta.

Para identificar as características dos casos com desvios maiores que 10% foram elaborados histogramas de ocorrência dos valores de cada parâmetro com os resultados dos 9.974 casos. A Figura 43 apresenta

histogramas dos casos com erros menores que -10% e com erros maiores que 10% para os parâmetros preditores do Metamodelo. Os histogramas em vermelho representam os casos com desvios negativos e os azuis com desvios positivos.





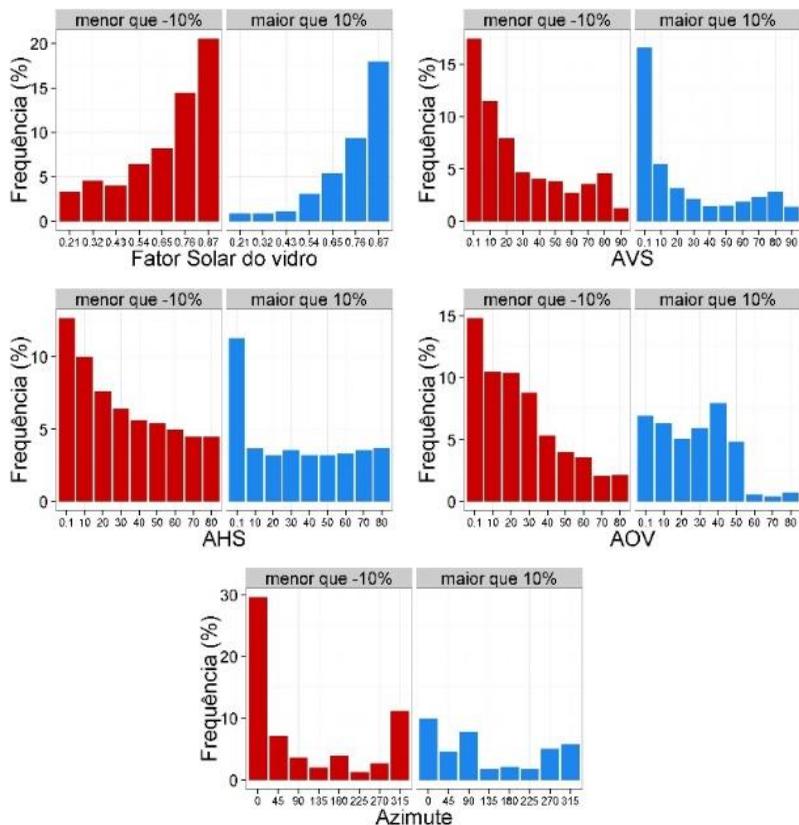


Figura 43. Histogramas da ocorrência de parâmetros em casos com desvios maiores que $5\text{kWh/m}^2\cdot\text{ano}$.

Segundo os histogramas da Figura 43, não ocorreu nenhum valor dos parâmetros preditores nos casos com erros percentuais maiores que 10%. Isso indica que todos os parâmetros podem apresentar casos com erros. Entretanto, para determinados parâmetros, os valores ocorrem com maior frequência que outros, sugerindo que estes valores causam mais erros. Os erros negativos são mais frequentes que os erros positivos, por isso que nos histogramas os erros menores que 10% somam mais casos que os erros maiores que 10%. Além disso, a ocorrência de valores que

causam maiores erros negativos são diferentes dos que causam erros positivos, exigindo uma avaliação separada destas duas situações.

Para os erros negativos, o horário de ocupação acarreta mais erros para ocupações de 24 horas. As propriedades térmicas da parede apresentam mais casos com erros com transmitâncias acima de $2,7\text{W}/\text{m}^2\cdot\text{K}$ e absorvância de 0,2. Os casos com massa térmica interna com menos de 20m^2 apresentam menos ocorrência de erros. E a exposição do piso e cobertura acarretam mais erros quando são adiabáticos. A altura da zona térmica, com o pé direito de 5 metros apresenta mais erros que os casos com as outras alturas, sejam elas maiores ou menores que isto. Da área envidraçada, observa-se que quanto maior o WWR ou o Fator Solar do vidro, maiores são as ocorrências de erros e o inverso ocorre para a transmitância da janela, onde quanto maior a transmitância, menor o erro. Para os três tipos de ângulos de sombreamento, quanto maior a obstrução, menor a ocorrência de erros. E quanto à orientação solar, os erros se concentram principalmente na orientação norte, quando o azimute é zero. Não existe uma tendência para se definir que quanto maior a possibilidade de trocas de calor interna e externa, a ocorrência de erros são maiores ou menores. Para casos como a exposição do piso e cobertura, a condição adiabática anula as trocas com o exterior e a ocorrência de erros aumenta, enquanto que para a transmitância da parede, o WWR e o fator solar do vidro, as condições de maior troca de calor acarretam em erros maiores.

Para os erros positivos, as diferenças de ocorrência de erros são menos evidentes. A exposição da cobertura e do piso como adiabática também acarreta em mais casos com erros maiores. Do mesmo modo, a área envidraçada e o fator solar do vidro também são diretamente proporcionais à ocorrência de erros. Porém, para os ângulos de

sombreamento, somente os casos sem obstrução vertical e horizontal se destacam com erros mais frequentes. E o ângulo de obstrução vizinha com obstruções de 60, 70 e 80 graus quase não possuem casos com erros positivos. O pé-direito acarreta em maiores erros para os casos com mais de 5 metros de altura, apresentando uma baixa ocorrência de erros para os casos com alturas menores que isto.

A ocorrência de casos com erros negativos ou positivos não evidencia nenhuma tendência, principalmente por não afetar a ocorrência de casos com erros menores que 10%. Os erros menores que 10% representam 99,2% dos casos, onde a ocorrência de todos os valores de parâmetros é quase homogênea por se tratar de uma amostra de 1,29 milhões de casos selecionados pela técnica de hipercubo latino. Desta forma, não há como identificar qual parâmetro, ou combinações de parâmetros, apresenta maior probabilidade de apresentar erros maiores. Contudo, os erros maiores ocorridos para 0,8% dos casos são intrínsecos ao processo de modelagem, que carrega erros da amostragem e do processo estatístico.

5.3. CONSIDERAÇÕES FINAIS

Neste capítulo o Metamodelo de Rede Neural Artificial foi explorado quanto a sua complexidade, avaliando-se o impacto do tamanho da amostra separada para treinamento e o tamanho da camada interna da arquitetura da Rede Neural Artificial. Foi definida a arquitetura com melhor desempenho em relação ao coeficiente de determinação (R^2), raiz do erro médio quadrático (RMSE) e o tempo computacional para treinamento. Além disso, foram feitos estudos de valores não vistos para

avaliar o erro de interpolação e extrapolação dos valores dos parâmetros preditores do Metamodelo. E por fim, para a arquitetura selecionada, foram analisados os erros da predição do Metamodelo para toda base de simulações com 1,29 milhões de casos. Com base nos resultados deste capítulo é possível destacar as seguintes considerações:

- a. O tamanho da amostra separada para treinamento é inversamente proporcional aos erros máximos estimados pelo método de dimensionamento pela média populacional. Este erro, em relação ao desvio padrão da população, variou de 9,15 a 0,29%, sendo que amostras com menos de 0,4% da população apresentaram erros amostrais maiores que 5%;
- b. As arquiteturas treinadas com amostras de até 0,3% não apresentaram estabilidade satisfatória pelo método de validação cruzada. Portanto, para arquiteturas de Redes Neurais Artificiais para resolução do problema de regressão deste trabalho, devem-se adotar amostras com mais de 0,3% da população, ou seja, mais de 3.380 casos;
- c. O aumento da camada interna tem maior influência no desempenho da Rede Neural do que o aumento da amostra. Ou seja, é melhor aumentar o número de neurônios na camada interna do que aumentar o número de casos simulados para elaboração do Metamodelo;
- d. A maior complexidade da arquitetura da Rede Neural Artificial acarreta em maior tempo de processamento

computacional, entretanto, o aumento desta complexidade não aumenta o desempenho do Metamodelo linearmente. Portanto, a análise do desempenho da predição deve levar em conta o uso de recursos computacionais em relação aos ganhos em precisão;

- e. As Redes Neurais Artificiais para predição de casos não vistos em seu treinamento apresentaram melhor desempenho para valores interpolados do que para valores extrapolados. Mesmo com a grande variabilidade de valores e parâmetros, com a garantia de homogeneidade da amostra pelo método de amostragem por hipercubo latino, a extrapolação de valores para predição de resultados não é recomendada;
- f. A Rede Neural Artificial treinada com amostra de 1% e com 72 nós na camada interna apresentou o melhor desempenho global para predição de todos os valores da base de simulações, com valores vistos, e para os casos com valores não vistos interpolados. Foi selecionada como Metamodelo final, após seus erros serem analisados;
- g. Para predição de toda base de dados simulados, o metamodelo final apresentou erros absolutos menores que $5\text{kWh/m}^2\cdot\text{ano}$ para 94,5% dos casos. Em relação a erros relativos, apenas 0,8% dos casos apresentam erros maiores que 10%. E para estes casos, não há um

parâmetro específico, ou combinações de parâmetros, que acarrete nos erros de predição do Metamodelo.

6. CONCLUSÕES

Neste último capítulo da tese são apresentadas as principais conclusões alcançadas, as limitações do trabalho e sugestões para trabalhos futuros.

Nesta tese, a hipótese de que um metamodelo pode representar com precisão o desempenho energético de zonas térmicas de edificações comerciais condicionadas artificialmente com diferentes complexidades de geometria, orientação, envoltória e cargas térmicas internas foram evidenciadas, principalmente, nos capítulos 4 e 5. Os resultados da comparação dos metamodelos com diferentes técnicas estatísticas de predição e com o refinamento do metamodelo de rede neural artificial para a base de 1,29 milhões de casos simulados validam a hipótese. A avaliação dos erros no capítulo 5 demonstra como o metamodelo final reproduz a variação de zonas térmicas de uma edificação comercial.

Mesmo com as considerações finais apontadas no final dos capítulos anteriores, ainda é possível destacar conclusões gerais importantes para colaboração ao conhecimento da área de desenvolvimento de metamodelos para predição de desempenho energético de edificações comerciais condicionadas artificialmente.

No processo de construção de um universo de casos de zonas térmicas de edifícios comerciais, através da parametrização, amostragem e simulações em cluster, se mostrou como um método automatizado pode ser desenvolvido para construção de bases de dados para diferentes finalidades. Podendo ser reproduzido para outros climas e outros parâmetros de interesse para análises de edificações. A visualização da base de dados permite estratificar parâmetros para análises específicas de tendências e correlações. Quando os resultados das simulações

evidenciaram a inter-relação entre as variáveis, pois para qualquer valor de parâmetro apresentou-se um intervalo significativo de resultados de carga anual de refrigeração. Nenhum parâmetro pode ser avaliado individualmente, sem sua interação com os demais parâmetros da edificação.

Na comparação das diferentes técnicas estatísticas de predição para elaboração dos metamodelos, a demanda de tempo de processamento destacou-se como variável importante. As técnicas estatísticas de predição apresentaram tendência linear entre os resultados simulados e preditos, apresentando médias equivalentes aos casos simulados e desvios padrões não maiores que 25%. Entretanto, o metamodelo de rede neural artificial apresentou o melhor desempenho entre os metamodelos treinados.

O metamodelo de Rede Neural Artificial foi explorado quanto a sua complexidade, avaliando-se o impacto do tamanho da amostra separada para treinamento e o tamanho da camada interna da arquitetura da Rede Neural Artificial. Conforme os resultados, o tamanho da amostra implica em um erro amostral relacionado a média e desvio padrão populacional, ou seja, quanto maior a amostra, melhor ela representa sua população. Portanto, o dimensionamento da amostra separada para treino é um passo para garantir a precisão do metamodelo. Nos processos de validação cruzada, as redes neurais artificiais treinadas com amostras de até 0,3% da base de dados, não apresentaram estabilidade satisfatória. Portanto, para arquiteturas de Redes Neurais Artificiais, para resolução do problema de regressão deste trabalho, devem-se adotar amostras com mais de 3.380 casos.

O aumento da camada interna da rede neural artificial apresentou maior influência do que o aumento da amostra. Portanto, é melhor aumentar o número de nós na camada interna do que aumentar o número de casos simulados para elaboração do Metamodelo. A maior complexidade da arquitetura da Rede Neural Artificial acarreta em maior tempo de processamento computacional, entretanto, o aumento desta complexidade não aumenta o desempenho do metamodelo linearmente.

Para o caso de estudo desta tese, a Rede Neural Artificial treinada com amostra de 1% da base de dados, aproximadamente 13 mil casos, e com 72 nós na camada interna apresentou o melhor desempenho global. Sendo selecionada como Metamodelo final. A análise dos erros deste metamodelo demonstrou que, para predição dos 1,29 milhões de casos da base de dados, ele apresenta precisão maior que 90% para predição de 99,2% dos casos.

6.1. LIMITAÇÕES DO TRABALHO

O metamodelo apresenta as mesmas limitações que o modelo de simulação que pretende substituir. Portanto, os metamodelos desenvolvidos nesta tese estão limitados às abrangências das simulações adotadas no EnergyPlus. A aplicação dos metamodelos deve se restringir às mesmas condições das simulação que resultaram na base de dados.

Quanto aos parâmetros construtivos, há a limitação das construções das paredes e coberturas que foram modeladas como objetos de apenas duas camadas, concreto e EPS. Estas construções são simplificações do comportamento da transmitância térmica e capacidade térmica de elementos mais comuns na construção civil. As paredes internas também

foram representadas por uma simplificação que representa a sua área equivalente em capacidade térmica de uma parede de concreto. O tipo de modelo para simulação do vidro não permite representa bem o comportamento de vidros insulados com superfícies de baixa emissividade. O percentual envidraçado da janela é representado por uma abertura no centro geométrico da parede, não podendo representar a distribuição da radiação refletida entre as paredes internas. Os elementos de sombreamento foram modelados como superfícies planas sem refletância, não representado o comportamento de elementos reflexivos ou de geometrias diferentes.

Os pisos e tetos entre pavimentos foram considerados adiabáticos, considerando que as condições de zonas térmicas de pavimentos tipos intermediários são iguais, ignorando a diferença entre alturas.

A carga térmica interna de equipamentos e iluminação foi configurada como uma fonte específica de calor, com frações de radiação e convecção pré-estabelecidas. Limitando a representação de equipamentos e iluminação com frações de ganhos de calor diferentes.

A infiltração de ar foi modelada como constante, e não foi modelada renovação de ar necessária para a ocupação de pessoas.

Os resultados de simulações apresentadas neste trabalho correspondem apenas a zonas térmicas perimetrais de 4,5m de profundidade, não podendo ser parâmetro para zonas internas.

Nos processos computacionais de modelagem estatística foram selecionados pacotes específicos para cada processo, utilizando funcionalidades do pacote caret para que todos os modelos fossem compatíveis. Entretanto, cada processo estatístico poderia ser configurado

com diversos parâmetros, onde diferentes configurações podem impactam para melhores resultados de cada método estatístico.

6.2. SUGESTÕES PARA TRABALHOS FUTUROS

Com base nos resultados desta tese, uma série de sugestões para trabalhos futuros pode ser oferecida.

Na elaboração da base de dados, mais parâmetros podem ser considerados. Os sombreamentos das janelas podem abranger mais tipos de elementos, além dos que promovem um ângulo de sombreamento vertical ou horizontal. Pode-se adotar elementos perfurados, telas, materiais reflexivos, etc. O sombreamento da vizinhança pode levar em conta sua reflexão, para além do efeito de sombreamento, também representar o efeito de reflexão da radiação solar pelas edificações vizinhas. Além de explorar o impacto de parâmetros de ventilação e iluminação natural, não abordados nesta tese.

Na comparação de diferentes metamodelos podem ser explorados outros tipos de técnicas estatísticas para predição. Neste trabalho foram adotadas configurações padrão para elaboração dos modelos estatísticos, cada modelo pode ser configurado para buscar melhores desempenhos.

Como variável dependente sugere-se elaborar metamodelos que reproduzam as cargas de refrigeração e aquecimento mensais, assim como seus valores de pico. As cargas de aquecimento podem ser úteis para climas mais frios. A predição de valores mensais serão úteis para avaliação do perfil de carga no decorrer do ano e os valores de pico para dimensionamento de sistemas de condicionamento de ar. No caso de redes neurais artificiais, a arquitetura deve ser ajustada para o maior número de

nós de saída, lembrando que quanto mais nós de entrada e saída, maior deve ser a camada interna.

Também é recomendado que em trabalhos futuros sejam avaliadas as questões qualitativas na aplicação de modelos, avaliando a experiência de usuários em diferentes complexidades de edificações na busca de uma metodologia amigável para avaliação do desempenho de edificações comerciais.

REFERÊNCIAS BIBLIOGRÁFICAS

AGUIRRE, Luis Antonio. **Introdução à identificação de sistemas**. Terceira ed. Belo Horizonte: Editora UFMG, 2007. p. 236

AL GHARABLY, Maged; DECAROLIS, Joseph F.; RANJITHAN, S. Ranji. An enhanced linear regression-based building energy model (LRBEM+) for early design. **Journal of Building Performance Simulation**, v.8. Fev. 2015.

AMORIM, Cláudia Naves David; CINTRA, Milena Sampaio; FREDERICO, Caio; TEIXEIRA, Júlia; SUDBRACK, Larissa Olivier. Energy efficiency code in Brazil: Experiences in the first public building labeled in Brasilia. **Proceedings...New York City: SimBuild 2010**. 2010

ASHRAE. **Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs**. Atlanta. American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2004.

ASHRAE. **Standard 90.1-2013 - Energy Standard for Buildings Except Low-Rise Residential Buildings**. Atlanta. American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2013.

AUTODESK. **Autodesk Ecotect Analysis**, 2011. Disponível em: <<http://usa.autodesk.com/ecotect-analysis/>>. Acesso em: 9 fev. 2015.

AUTODESK. **Autodesk Vasari**, 2014. Disponível em: <<http://autodeskvasari.com/>>. Acesso em: 9 fev. 2015.

AUTODESK. **Project EnergyPlus Cloud**. . Disponível em: <<https://beta.autodesk.com/callout/?callid=5366FFEA6FCC4F6A94E19B78A62698A2>>. Acesso em: 9 fev. 2015.

AYDINALP, Merih; UGURSAL, V.I.; FUNG, Alan S. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. **Applied Energy**, v. 71, n. 2, p. 87–110, fev. 2002.

BACKHAUS, Julia; TIGCHELAAR, Casper; BEST-WALDHOBER, Marjolein de. **Key findings & policy recommendations to improve effectiveness of Energy Performance Certificates & the Energy Performance of Buildings Directive. Summary of IDEAL EPBD research findings and policy recommendations.** [s.l]. 2011.

BORGSTEIN, Edward H.; LAMBERTS, Roberto. Developing energy consumption benchmarks for buildings: Bank branches in Brazil. **Energy and Buildings**, v. 82, p. 82–91, out. 2014.

BREIMAN, L. Random forests. **Machine Learning**, v.45. p. 5–32, Out. 2001.

BREIMAN, L. CUTLER, A. **Breiman and Cutler's random forests for classification and regression.** R Package: 'randomForest' version 4.6-10. 2014.

CARLO, Joyce. **Desenvolvimento de Metodologia de Avaliação da Eficiência Energética do Envolvimento de Edificações Não-residenciais.** 2008. 193 f. Tese (Doutorado em Engenharia Civil)—Programa de Pós Graduação em Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2008.

CARLO, Joyce; LAMBERTS, Roberto. **Processamento de arquivos climáticos para simulação do desempenho energético de edificações.** Florianópolis. 2005. Disponível em: <http://www.labee.ufsc.br/sites/default/files/arquivos_climaticos/RT200504.pdf>.

CARLO, Joyce; LAMBERTS, Roberto. Development of envelope efficiency labels for commercial buildings: Effect of different variables on electricity consumption. **Energy and Buildings**, v. 40, n. 11, p. 2002–2008, jan. 2008.

CASTRO, Luíza Tavares de; WESTPHAL, Fernando Simon. Comparativo entre o método prescritivo e simulação para análise de desempenho de envoltória. In: Encontro Nacional de Conforto no Ambiente Construído. 2011. **Anais...Búzios: ENCAC.** 2011

CATALINA, Tiberiu; IORDACHE, Vlad; CARACALEANU, Bogdan. Multiple regression model for fast prediction of the heating energy demand. **Energy and Buildings**, v. 57, p. 302–312, fev. 2013.

CHOU, Jui Sheng; BUI, Dac Khuong. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. **Energy and Buildings**, v. 82, p. 437–446, jul. 2014.

CLARKE, J.A. **Energy Simulation in Building Desing**. Second ed. Oxford: Butterworth-Heinemann, 2001. p. 362

CÓSTOLA, Daniel. **External coupling of building energy simulation and building element heat, air and moisture simulation**. 2011. 134 f. Thesis (PhD-Thesis)—Eindhoven University of Technology, Eindhoven. 2011.

CRAWLEY, Drury; HAND, J.W.; KUMMERT, M.; GRIFFITH, Brent T. Contrasting the capabilities of building energy performance simulation programs. In: International IBPSA Building Simulation Conference. **Proceedings...**Montréal: BS2005. 2005.

CRAWLEY, Drury; LAWRIE, L.K. EnergyPlus: creating a new-generation building energy simulation program. **Energy and Buildings**, v. 33, n. 4, p. 319–331, 2001.

CRAWLEY, Drury; SANDER, Daniel M.; CORNICK, Steve; NEWSHAM, Guy R. Development of a simple model to relate heating and cooling energy to building envelope thermal characteristics. In: International IBPSA Building Simulation Conference. **Proceedings...** Adelaide, Australia: BS1993. 1993.

DATTA, D; TASSOU, S.A.; MARRIOTT, D. Application of neural networks for the prediction of the energy consumption in a supermarket. In: CLIMA 2000 Conference. **Anais....**Brussels: CLIMA2000. 1997. p98.

DE WILDE, Pieter. The gap between predicted and measured energy performance of buildings: A framework for investigation. **Automation in Construction**, v. 41, p. 40–49, maio 2014.

DE WILDE, Pieter; TIAN, Wei; AUGENBROE, Godfried. Longitudinal prediction of the operational energy use of buildings. **Building and Environment**, v. 46, n. 8, p. 1670–1680, 2011.

EKICI, Betül Bektas; AKSOY, U Teoman. Prediction of building energy consumption by using artificial neural networks. **Advances in Engineering Software**, v. 40, n. 5, p. 356–362, 2009.

EKONOMOU, L. Greek long-term energy consumption prediction using artificial neural networks. **Energy**, v. 35, n. 2, p. 512–517, 2010.

ELDRED, M. S.; GIUNTA, A. A.; BROWN, S. L.; ADAMS, B. M.; DUNLAVY, D. M.; EDDY, J. P.; GAY, D. M.; GRIFFIN, J. D.; HART, W. E.; HOUGH, P. D.; KOLDA, T. G.; MARTINEZ-CANALES, M. L.; SWILER, L. P.; WATSON, J.; WILLIAMS, P. J. **DAKOTA, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis**. Livermore, 2013.

ENERGYPLUS. **Getting Started with EnergyPlus**. [s.l.] 2013. p. 68

ENERGYPLUS. **Auxiliary EnergyPlus Programs: Extra programs for EnergyPlus**. [s.l.] US Department of Energy, 2013b. p. 222

FONSECA, Raphaela Walger da; DIDONÉ, Evelise Leite; PEREIRA, Fernando Oscar Ruttkay. Using artificial neural networks to predict the impact of daylighting on building final electric energy requirements. **Energy and Buildings**, v. 61, p. 31–38, jun. 2013.

FONTENELLE, Marília Ramalho; SANTOS, Maria Júlia De Oliveira. Análise da aplicação de parâmetros da envoltória no método prescritivo do RTQ-C: A relação do bioclimatismo com a eficiência energética de edifícios. In: Encontro Nacional de Conforto no Ambiente Construído. 2011. **Anais...** Búzios: ENCAC 2011. 2011.

FORRESTER, Alexander I.J.; SOBESTER, Andrés; KEANE, Andy J. **Engineering design via surrogate modelling: a practical guide**. Chichester: John Wiley & Sons, Ltd, 2008. p. 238

FOSSATI, Michele; LAMBERTS, Roberto. Eficiência energética da envoltória de edifícios de escritórios de Florianópolis: discussões sobre a aplicação do método prescritivo do RTQ-C. **Ambiente Construído**, v. 10, n. 2, p. 59–69, 2010.

FOUCQUIER, Aurélie; ROBERT, Sylvain; SUARD, Frédéric; STÉPHAN, Louis; JAY, Arnaud. State of the art in building modelling and energy performances prediction: A review. **Renewable and Sustainable Energy Reviews**, v. 23, p. 272–288, jul. 2013.

FRIEDMAN, J.H. Multivariate Adaptive Regression Splines. **Annals of Statistics**, v.19, n.1, p.1-141. 1991.

FUMO, Nelson. A review on the basics of building energy estimation. **Renewable and Sustainable Energy Reviews**, v. 31, p. 53–60, mar. 2014.

GIARETTA, Rafael Fernando; TEIXEIRA, Cristiano André; WESTPHAL, Fernando Simon. Processo de etiquetagem Procel: Dificuldades encontradas na aplicação do método prescritivo. In: Encontro Nacional de Tecnologia no Ambiente Construído. 2012. **Anais...** Juiz de Fora: ENTAC2012. 2012.

GORISSEN, Dirk; COUCKUYT, I. A surrogate modeling and adaptive sampling toolbox for computer based design. **The Journal of Machine Learning Research**, v. 11, p. 2051–2055, 2010.

GOULART, Solange; LAMBERTS, Roberto; FIRMINO, Samanta. **Dados climáticos para projeto e avaliação energética de edificações para 14 cidades brasileiras**. Florianópolis: Núcleo de Pesquisa em Construção/UFSC. 1998.

GOWRI, Krishnan. **Cloud-based Energy Simulation using EnergyPlus**. Autodesk. 2015. Disponível em: <<https://global.gotowebinar.com/join/over.tmpl?webinar=3387791401985993474>>. Acesso em: 10 fev. 2015

HAYKIN, Simon. **Redes Neurais. Princípios e Prática**. Segunda ed. São Paulo: Bookman Companhia Editora. 2001. p. 899

HENSEM, Jan L.M.; LAMBERTS, Roberto. Introduction to building performance simulation. In: **Building performance simulation for design and operation**. New York: Spon Press, 2011. p. 1–14.

HEO, Yeonsook; GRAZIANO, Diane; GUZOWSKI, Leah; MUEHLEISEN, Ralph T. Evaluation of calibration efficacy under different levels of uncertainty. In: International IBPSA Building Simulation Conference. 2013. **Proceedings...** Chambéry, France: BS2013. 2013.

HOPFE, Christina Johanna; AUGENBROE, Godfried; HENSEN, Jan L.M. Multi-criteria decision making under uncertainty in building performance assessment. **Building and Environment**, v. 69, p. 81–90, nov. 2013.

HOPFE, Christina JOHANNA; Hensen, Jan L.M. Uncertainty analysis in building performance simulation for design support. **Energy and Buildings**, v. 43, n. 10, p. 2798–2805, out. 2011.

INMETRO. **RTQ-C: Regulamento Técnico da Qualidade do Nível de Eficiência Energética de Edifícios Comerciais, de Serviços e Públicos**. Brasília, 2009.

INMETRO. **RTQ-R: Regulamento Técnico da Qualidade do Nível de Eficiência Energética de Edifícios Residenciais**. Brasília, 2011.

INSTITUTE FOR MARKET TRANSFORMATION. **Building Rating**. Disponível em: <<http://buildingrating.org/>>. Acesso em: 24 mar. 2015.

JEPLUS. **JePlus User's Manual. Version 1.5**. Disponível em: <http://www.jeplus.org/wiki/doku.php?id=docs:manual_1_5>. Acesso em: 1 jan. 2014.

KALAY, Yehuda E. Performance-based design. **Automation in Construction**, v. 8, n. 4, p. 395–409, abr. 1999.

KALOGIROU, Soteris A. Applications of artificial neural-networks for energy systems. **Applied Energy**, v. 67, p. 17–35, 2000.

KALOGIROU, Soteris A.; BOJIC, Milorad. Artificial neural networks for the prediction of the energy consumption of a passive solar building. **Energy**, v. 25, n. 5, p. 479–491, 2000.

KARAMPELAS, Panagiotis; VITA, Vita; PAVLATOS, Christos; MLADENOV, Valeri; EKONOMOU, Lambros. Design of artificial neural network models for the prediction of the Hellenic energy consumption. In: Symposium on Neural Network Applications in Electrical Engineering. 2010. **Proceedings...** Belgrade: NEUREL 2010. 2010.

KARATASOU, S; SANTAMOURIS, Mat; GEROS, V. Modeling and predicting building's energy use with artificial neural networks: Methods and results. **Energy and Buildings**, v. 38, p. 949–958, 2006.

KARATZOGLOU, Alexandros; SMOLA, Alex; HORNIK, Kurt (2015). **kernlab: Kernel-based Machine Learning Lab**. R Package : 'kernlab' version 0.9-20.

KAVAKLIOGLU, Kadir. Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. **Applied Energy**, v. 88, n. 1, p. 368–375, jan. 2011.

KIM, Byungseon S.; DEGELMAN, Larry O. An interface system for computerized energy analyses for building designers. **Energy and Buildings**, v. 27, n. 1, p. 97–107, fev. 1998.

KIM, Hyunjoo; STUMPF, Annette; KIM, Wooyoung. Analysis of an energy efficient building design through data mining approach. **Automation in Construction**, v. 20, n. 1, p. 37–43, jan. 2011.

KINGRE, H.J. **Bezier Curve for Metamodeling of Simulation Output**. Thesis (Master Thesis). Louisiana State University. 2004.

KOROLJA, Ivan; MARJANOVIC-HALBURD, Ljiljana; ZHANG, Yi; HANBY, Victor I. UK office buildings archetypal model as methodological approach in development of regression models for predicting building energy consumption from heating and cooling demands. **Energy and Buildings**, v. 60, p. 152–162, maio 2013a.

KOROLIJA, Ivan; ZHANG, Yi; MARJANOVIC-HALBURD, Ljiljana; HANBY, Victor I. Regression models for predicting UK office building energy consumption from heating and cooling demands. **Energy and Buildings**, v. 59, p. 214–227, abr. 2013b.

KUHN, Max. **Caret: Classification and Regression Training**. R Package: ‘caret’ version 6.0-41. 2015.

KUHN, Max; JOHNSON, Kjell. **Applied Predictive Modeling**. New York: Springer, 2013. p. 595

KUMAR, Rajesh; AGGARWAL, R.K.; SHARMA, J.D. Energy analysis of a building using artificial neural network: A review. **Energy and Buildings**, v. 65, p. 352–358, out. 2013.

KUYUK, H S; KAYIKCI, R; ÖZKAN, O; BAYRAKTAROĞLU, S. Modelling Turkey’s Energy Consumption Based on Artificial Neural Network. In: Ninth International Conference on Machine Learning and Applications. 2010. **Proceedings...** Bethesda: ICMLA 2010. 2010

LABEEE. **Arquivos climáticos em formato TRY, SWERA, CSV e BIN**. Disponível em: <<http://www.labee.ufsc.br/downloads/arquivos-climaticos/formato-try-swera-csv-bin>>. Acesso em: 7 dez. 2012.

LAWRENCE, Steve; GILES, C.Lee; TSOI, Ah Chung. **What Size Neural Network Gives Optmail Generalization? Convergence Properties of Backpropagation**. Queensland, Australia. 1996.

LEIPZIGER, David. **Comparing Building Energy Performance Measurement. A framework for international energy efficiency assessment systems**. Washington, DC, 2013.

LI, Qiong; MENG, Qinglin; CAI, Jiejun; YOSHINO, Hiroshi; MOCHIDA, Akashi. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. **Energy Conversion and Management**, v. 50, n. 1, p. 90–96, jan. 2009.

LI, Zhengwei; HUANG, Gongsheng. Re-evaluation of building cooling load prediction models for use in humid subtropical area. **Energy and Buildings**, v. 62, p. 442–449, jul. 2013.

LIAO, Zaiyi; DEXTER, A.L. A simplified physical model for estimating the average air temperature in multi-zone heating systems. **Building and Environment**, v. 39, n. 9, p. 1013–1022, set. 2004.

LONG, Nicholas L; BALL, Brian L; FLEMING, Katherine A; MACUMBER, Daniel L. Scaling building energy modeling horizontally in the cloud with OpenStudio. In: ASHRAE/IBPSA-USA Building Simulation Conference. **Proceedings...Atlanta**. 2014

LOPES, Mauricio Nath; IWAMOTO, Gabriel; RIELLA, Humberto Leitão; LAMBERTS, Roberto; CAMPOS, Lucila Maria Souza. Development of Computational Tool to Evaluate the Building Energy Efficiency Level According to the Brazilian Labeling. In: International IBPSA Building Simulation Conference. 2011. **Proceedings...Sydney, Australia: BS2011**. 2011.

LORENA, Ana Carolina; CARVALHO, André C.P.L.F. DE. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.

MACDONALD, Iain A. Comparison of sampling techniques on the performance of Monte Carlo based sensitivity analysis. In: International IBPSA Building Simulation Conference. 2009. **Proceedings...Glasgow, Scotland: BS2009**. 2009.

MACKAY, D.J.C. Introduction to Gaussian processes. **NATO ASI Series of Computer and Systems Sciences**. 1998.

MAHDAVI, Ardeshir; EL-BELLAHY, Shukri. Effort and effectiveness considerations in computational design evaluation: a case study. **Building and Environment**, v. 40, p. 1651–1664, 2005.

MAILE, Tobias; BAZJANAC, Vladimir; FISCHER, Martin. A method to compare simulated and measured data to assess building energy performance. **Building and Environment**, v. 56, p. 241–251, 2012.

MALKAWI, Ali; WAEGEL, Alex. Rapid modeling of buildings with calibrated normative models. In: International IBPSA Building Simulation Conference . 2013. **Proceedings...**Chambéry, France: BS2013. 2013

MAO, Chunliu; HABERL, Jeff S; BALTAZAR, Juan-Carlos. **Literature Review on the History of Building Peak Load and Annual Energy Use Calculation Methods in the U.S.** Bryan, Texas, 2013a.

MAO, Chunliu; HABERL, Jeff S.; BALTAZAR, Juan-Carlos. Peak heating/cooling load design methods: How we got to where we are today in the U.S. In: International IBPSA Building Simulation Conference. 2013. **Proceedings...**Chambéry, France: BS2013. 2013b

MARSH, Andrew. **Performance Analysis and conceptual design.** Thesis. 1999. 216 f. (PhD-Thesis)—Sydney, Australia: University of Western Australia, 1999.

MELO, Ana Paula. **Desenvolvimento de um método para estimar o consumo de energia de edificações comerciais através da aplicação de redes neurais.** Tese. (Doutorado em Engenharia Civil)—Programa de Pós Graduação em Engenharia Civil. Florianópolis: Universidade Federal de Santa Catarina, 2012.

MELO, Ana Paula; CÓSTOLA, Daniel; LAMBERTS, Roberto; HENSEN, Jan L.M. Avaliação comparativa entre o método prescritivo e o método de simulação para determinação do nível de eficiência da envoltória de edificações comerciais. In: Encontro Nacional de Conforto no Ambiente Construído. 2011. **Anais...**Búzios: ENCAC 2011. 2011a

MELO, Ana Paula; CÓSTOLA, Daniel; LAMBERTS, Roberto; HENSEN, Jan L.M. Assessing the accuracy of a simplified building energy simulation model using BESTEST: The case study of Brazilian regulation. **Energy and Buildings**, v. 45, p. 219–228, fev. 2012.

MELO, Ana Paula; LAMBERTS, Roberto; CÓSTOLA, Daniel; HENSEN, Jan L.M. Capability and deficiency of the simplified model for energy calculation of commercial buildings in the Brazilian regulation. In: International IBPSA Building Simulation Conference. 2011. **Proceedings...**Sydney, Australia: BS2011. 2011b.

MELO, Ana Paula; LAMBERTS, Roberto; CÓSTOLA, Daniel; HENSEN, Jan L.M. Development of a Method to Predict Building Energy Consumption through an Artificial Neural Network Approach. In: International IBPSA Building Simulation Conference. 2013. **Proceedings...**Chambéry, France: BS2013. 2013.

MEYER, David; DIMITRIADOU, Evgenia; HORNİK, Kurt; WEINGESSEL, Andreas; Leisch, Friedrich; CHANG, Chih-Chung; LIN, Chih-Chen; TU, Wien. **Misc Functions of the Department of Statistics (e1071)**. R Package: ‘e1071’ version 1.6-4. 2014.

MILBORROW, S. **Multivariate Adaptive Regression Spline Models**. R Package: ‘earth’ version 3.2-7. 2014.

NABONI, Emanuele; ZHANG, Yi; MACCARINI, Alessandro; HIRSCH, Elian; LEZZI, Daniele. Extending the use of parametric simulation in practice through a cloud based online service. In: IBPSA Italy conference. 2013. **Proceedings...** Bozen-Bolzano: BSA2013. 2013.

NEW, Joshua; SANYAL, Jibonananda; BHANDARI, Mahabir; SHRESTHA, Som. Autotune e+ building energy models. In: 5th National Conference of IBPSA-USA. **Proceedings...**Madison: SimBuild 2012. 2012.

NREL. **OpenStudio: A DAKOTA DDACE Tutorial**. . Disponível em: <http://openstudio.nrel.gov/sites/openstudio.nrel.gov/files/SmOff_DDACE_tutorial.pdf>. Acesso em: 7 mar. 2013.

NREL. **OpenStudio**. National Renewable Energy Laboratory.[n.l.] 2014. . Disponível em: <<http://www.openstudio.net>>

OLIVEIRA, Helena S.; GUTIERREZ, Grace C.R. Análise do edifício SESC Canoas-RS para classificação do nível de eficiência energética pelo RTQ-C: Etiqueta parcial da envoltória. In: Encontro Nacional de Tecnologia no Ambiente Construído. **Anais...**Juiz de Fora: ENTAC 2012. 2012.

OLSEN, Erik; IVERSEN, Michael. **Overview of Simulation. Presentation for architects.** Illinois: IBPSA-USA. 144 slides. Apresentação em Power-point. 2011.

PACHECO, Giovanni Hudson; OLIVEIRA, Paolo Américo de; DIAS, Alice Rück Drummond; MARTINS, Micael Ribeiro; PEDRINI, Aldomar. Influência dos ângulos de sombreamento no resultado do indicador de consumo da envoltoria através do método prescritivo do RTQ-C. In: Encontro Nacional de Tecnologia no Ambiente Construído. **Anais...**Juiz de Fora: ENTAC 2012. 2012

PAPAMICHAEL, Konstantinos; PORTA, John La; CHAUVET, Hannah. Decision making through use of interoperable simulation software. In: International IBPSA Building Simulation Conference. 1997. **Proceedings...**Prague, Czech Republic:BS1997. 1997.

PEDRINI, Aldomar. **Integration of low energy strategies to the early stages of design process of office buildings in warm climate.** Thesis. 2003. 259 f. (PhD-Thesis)—Brisbane: The University of Queensland, 2003.

PEDRINI, Aldomar; LIMA, Glênio L F; OLIVEIRA, Paolo Américo de; TRINDADE, Sileno C. Análise comparativa da eficiência energética de envoltórias segundo o regulamento de etiquetagem de edifícios. In: Encontro Nacional de Tecnologia do Ambiente Construído. 2010. **Anais...**Canela: ENTAC2010. 2010

PÉREZ-LOMBARD, Luis; ORTIZ, José; GONZÁLEZ, Rocío; MAESTRE, Ismael R. A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes. **Energy and Buildings**, v. 41, n. 3, p. 272–278, mar. 2009.

PINA, Aline Aparecida. **Metodologias de análise, síntese e otimização de sistemas para produção de petróleo offshore através de metamodelos e exames de partículas**. Tese (Doutorado em Engenharia Civil). 2010. COPPE. Rio de Janeiro: Universidade Federal do Rio de Janeiro. 2010.

QIAN, Zhiguang; SEEPERSAD, Carolyn Conner; JOSEPH, V. Roshan; ALLEN, Janet K.; WU, C.F.Jeff. Building surrogate models based on detailed and approximate simulations. **Journal of Mechanical Design**, v. 128, n. 4, p. 668–667, 2006.

QIN, Rong; YAN, Da; ZHOU, Xin; JIANG, Yi. Research on a dynamic simulation method of atrium thermal environment based on neural network. **Building and Environment**, v. 50, p. 214–220, abr. 2012.

R, Developed Core Team. **R: A language and environment for statistical computing**, 2014. Disponível em: <<http://www.r-project.org>>

REICHARD, Georg; PAPAMICHAEL, Konstantinos. Decision-making through performance simulation and code compliance from the early schematic phases of building design. **Automation in construction**, v. 14, p. 173–180, 2005.

RIPLEY, B; VENABLES, W. **Feed-forward Neural Networks and Multinomial Log-Linear Models**. R Package: ‘NNET’ version 7.3-8. 2014.

RMI. **Merritt Jenkins on the History of Building Energy Modeling**. Rocky Mountain Institute, 2010. Disponível em: <<http://www.rmi.org/MerrittHistoryBEM>>. Acesso em 15 de abril de 2013.

RODRIGUES, Clara Olívio de Medeiros; DIAS, Alice Rück Drummond; OLIVEIRA, Paolo Américo de; HAZBOUN, Viviane Diniz; MARTINS, Micael Ribeiro; PACHECO, Giovani Hudson; CARMO, Fernando Augusto Leite Cortez do; PEDRINI, Aldomar. Aplicação da etiquetagem do nível de eficiência energética de edifícios (505). In: Encontro Nacional de Tecnologia do Ambiente Construído. 2012. **Anais...Juiz de Fora: ENTAC2012.** 2012.

SANTAMOURIS, Mat. **Energy performance of residential buildings: a practical guide for energy rating and efficiency.** London: Routledge, 2005. p. 152

SARLE, Warren S. **comp.ai.neural-nets.** Disponível em: <<ftp://ftp.sas.com/pub/neural/FAQ.html>>. Acesso em: 17 fev. 2015.

SIGNOR, Régis. **Análise de regressão do consumo de energia elétrica frente a variáveis arquitetônicas para edifícios comerciais climatizados em 14 capitais brasileiras.** Dissertação. 1999. 314 f. (Mestrado em Engenharia Civil)—Florianópolis: Universidade Federal de Santa Catarina, 1999.

SIGNOR, Régis; WESTPHAL, Fernando Simon; LAMBERTS, Roberto. Regression analysis of electric energy consumption and architectural variables of conditioned commercial buildings in 14 Brazilian cities. In: International IBPSA Building Simulation Conference. 2001. **Anais...** Rio de Janeiro: BS2001. 2001.

SILVA, Arthur Santos; GHISI, Enedir. Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation. **Energy and Buildings**, v. 76, p. 381–391, 2014.

SOUZA, Clarice Bleil de; KNIGHT, Ian. Thermal performance simulation from an architectural design viewpoint. In: International IBPSA Building Simulation Conference. 2007. **Anais...**Beijing, China: BS2007. 2007.

STEPHENS, D.W.; GORISSEN, Dirk; DHAENE, T. Surrogate based sensitivity analysis of process equipment. In: Seventh International Conference on CFD in the Minerals and Process Industries. **Proceedings...** Trondheim, Norway: CFD2011. 2011.

TSANAS, Athanasios; XIFARA, Angeliki. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. **Energy and Buildings**, v. 49, p. 560–567, jun. 2012.

VAN DAM, E. R.; HUSSLAG, B.; DEN HERTOOG, D.; MELISSEN, H. Maximin latin hypercube designs in two dimensions. **Operations Research**, v. 55, p. 158–169, 2007.

VANGIMALLA, Prasanthi R.; J.OLBINA, Svetlana; R.ISSA, Raymond; HINZE, Jimmie. Validation of Autodesk Ecotect Accuracy for thermal and daylighting simulations. In:2011 Winter Simulation Conference. **Proceedings...**Arizona:WSC2011. 2011

VENABLES, W N; RIPLEY, B D. **MASS: modern applied statistics with issues of accuracy and scale**. R Package: 'MASS'. 2002.

VERSAGE, Rogério; LAMBERTS, Roberto. **Equações prescritivas para o regulamento de etiquetagem de eficiência energética de edificações residenciais**. Florianópolis. 2011.

WESTPHAL, Fernando Simon; LAMBERTS, Roberto. Regression analysis of electric energy consumption of commercial buildings in Brazil. In: International IBPSA Building Simulation Conference. 2007. **Proceedings...**Beijing, China: BS2007. 2007.

YAMAKAWA, Mary A.; WESTPHAL, Fernando Simon. Influência do percentual de abertura nas fachadas e do fator solar dos vidros na etiquetagem do Procel/INMETRO: Método prescritivo x simulação. In: Encontro Nacional de Conforto no Ambiente Construído. 2011. **Anais...** Búzios: ENCAC 2011. 2011

YANG, Jin; RIVARD, Hugues; ZMEUREANU, Radu. Building energy prediction with adaptive artificial neural networks. In: International IBPSA Building Simulation Conference. 2005. **Proceedings...** Montréal, Canada:BS2005. 2005.

ZHANG, Yi. “Parallel” EnergyPlus and the development of a parametric analysis tool. In: International IBPSA Building Simulation Conference. 2009. **Proceedings...** Glasgow, Scotland: BS2009. 2009.

ZHANG, Yi; KOROLIJA, Ivan. Performing complex parametric simulations with jEPlus. In: International Conference on Sustainable Energy Technologies. 2010. **Proceedings...** Shanghai, China: SET2010. 2010.

ZHAO, Hai Xiang; MAGOULÈS, Frédéric. A review on the prediction of building energy consumption. **Renewable and Sustainable Energy Reviews**, v. 16, n. 6, p. 3586–3592, 2012.

ZHOU, Qiang; WANG, Shengwei; XU, Xinhua; XIAO, Fu. A grey-box model of next-day building thermal load prediction for energy-efficient control. **International Journal of Energy Research**, v. 32, p. 1418–1431, Jul. 2008.