

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

**MODELAGEM DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA
NORMAL EM BOTUCATU: APRENDIZADO DE MÁQUINA,
ESTATÍSTICO E LINKE**

CÍCERO MANOEL DOS SANTOS

Tese apresentada à Faculdade de
Ciências Agronômicas da UNESP –
Câmpus de Botucatu, para obtenção do
título de Doutor em Agronomia
(Irrigação e Drenagem).

BOTUCATU-SP
MARÇO – 2016

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

**MODELAGEM DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA
NORMAL EM BOTUCATU: APRENDIZADO DE MÁQUINA,
ESTATÍSTICO E LINKE**

CÍCERO MANOEL DOS SANTOS

Orientador: Prof. Dr. João Francisco Escobedo

Tese apresentada à Faculdade de
Ciências Agronômicas da UNESP –
Câmpus de Botucatu, para obtenção do
título de Doutor em Agronomia
(Irrigação e Drenagem).

BOTUCATU-SP
MARÇO – 2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO - DIRETORIA TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - UNESP - FCA - LAGEADO - BOTUCATU (SP)

S237m Santos, Cícero Manoel dos, 1987-
Modelagem da irradiação direta na incidência normal em Botucatu: aprendizado de máquina, estatístico e linke / Cícero Manoel dos Santos. - Botucatu : [s.n.], 2016
xiv, 115 f. : fots. color.; grafs. color., ils. color., tabs.

Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrônômicas, Botucatu, 2016

Orientador: João Francisco Escobedo

Inclui bibliografia

1. Redes neurais (Computação). 2. Mudanças climáticas. 3. Radiação solar. 4. Energia solar. I. Escobedo, João Francisco. II. Universidade Estadual Paulista "Júlio de Mesquita Filho" (Câmpus de Botucatu). Faculdade de Ciências Agrônômicas. III. Título.

CERTIFICADO DE APROVAÇÃO

**TÍTULO DA TESE: MODELAGEM DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL EM
BOTUCATU: APRENDIZADO DE MÁQUINA, ESTATÍSTICO E LINKE**

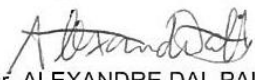
AUTOR: CÍCERO MANOEL DOS SANTOS

ORIENTADOR: JOÃO FRANCISCO ESCOBEDO

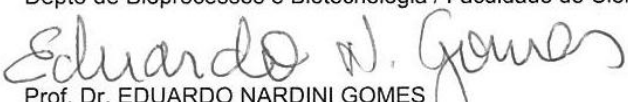
Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA
(IRRIGAÇÃO E DRENAGEM), pela Comissão Examinadora:



Prof. Dr. JOÃO FRANCISCO ESCOBEDO
Depto de Engenharia Rural / Faculdade de Ciências Agronômicas de Botucatu



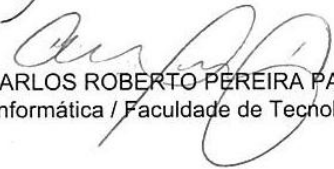
Prof. Dr. ALEXANDRE DAL PAI
Depto de Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônômicas de Botucatu



Prof. Dr. EDUARDO NARDINI GOMES
Depto de Engenharia de Pesca / UNESP - Câmpus Experimental de Registro



Profa. Dra. SILVIA HELENA M. GORLA DA SILVA
Depto de Engenharia e Pesca / UNESP - Câmpus Experimental de Registro



Prof. Dr. CARLOS ROBERTO PEREIRA PADOVANI
Depto de Informática / Faculdade de Tecnologia de Botucatu

Botucatu, 04 de março de 2016.

OFEREÇO

A DEUS por sua perfeição em toda criação. Por conceder o Dom da vida... Por permitir a consolidação de mais uma fase em minha vida.

A Santa e Virgem Maria pelo “colo” concebido e sempre está ao meu lado intercedendo nos momentos bons e nos mais difíceis da minha vida.

DEDICO

A minha mãe Josefa e ao meu pai José Cândido (em memória), pelas lições e aprendizados na vida. Que, apesar das dificuldades, sempre me educaram e direcionaram para os melhores caminhos possíveis a trilhar. Foram pelos “corretivos” na adolescência que pude dar uma nova direção a minha vida. Obrigado por tudo!

Aos meus irmãos (Nado, Gil e Flavinho) e as irmãs (Rosa, Ninha, Neta e Nelminha) por sempre estarem ao meu lado nos momentos bons e difíceis da vida. Pelos conselhos, pelo convívio, pelas brincadeiras... Obrigado!

A minha sobrinha (afilhada) Rafaela pelos poucos, mas carinhosos momentos de brincadeiras e de raiva!

A minha cunhada (Novinha) e ao meu cunhado (Cícero), pelo apoio nos momentos de dificuldades.

AGRADECIMENTOS

Ao Prof.^o Dr. João Francisco Escobedo pela orientação, pelo conhecimento, pelo ensinamento, pelos conselhos, pela paciência e pela amizade durante o curso de doutorado, sem os quais nada disso teria sido realizado. Meus sinceros agradecimentos!

A Universidade Estadual Paulista, a Faculdade de Ciências Agronômicas e ao Programa de Pós-Graduação em Irrigação e Drenagem (UNESP/FCA/BOTUCATU) pelo comprometimento em oferecer uma ótima estrutura e formar para o mercado de trabalho excelente Profissionais, Mestres e Doutores. Meus singelos agradecimentos por possibilitar a conclusão de mais uma etapa em minha vida profissional.

A todos os professores do Departamento de Engenharia que contribuíram diretamente ou indiretamente para minha formação. Muito obrigado!

Aos professores: Dr. Alexandre Dal Pai, Dr. Eduardo Nardini, Dr^a. Silvia Helena e Dr. Carlos Padovani, por disponibilizarem seu tempo para compor a banca avaliadora da tese. Muito obrigado!

A Prof^a. Dr^a. Silvia Helena Modenese Gorla da Silva pelo compartilhamento do conhecimento das técnicas de aprendizado de máquinas e a disposição de seu tempo. Meus sinceros agradecimentos!

Aos amigos e professores: Dr. Alexandre Dal Pai, Dr. Érico Tadao Teramoto, Dr. Enzo Dal Pai e Dr. Eduardo Nardini Gomes, pelos ensinamentos, pelos conselhos, pela amizade, pela contribuição direta e indiretamente para este trabalho. Obrigado!

Aos amigos de laboratório: Maurício, Esteban, Lucas, Daniel e Bartolomeu, pelos conhecimentos e “resenhas” compartilhados durante o doutorado.

Aos amigos e funcionários do departamento: Taíza, Maria, Raquel... Pela amizade, pelos ensinamentos e conselhos.

A todos colegas e amigos de turma do curso de Doutorado, pelo companheirismo e momentos de descontrações.

Aos amigos: Cícero Teixeira e Rafael Pombo meus sinceros agradecimentos.

Ao GPP (Grupo de partilhas de Profissionais) e todos os participantes (Dila, Maria, Jack, Cícero, Patrícia, Maíla, Tiago, Jorge, Paula, Meirinha... e todos os demais) Vocês foram (serão) muito importante em minha vida e estadia em Botucatu. Meu sincero e fraterno OBRIGADO!

Ao Grupo de Oração Jovem Cenáculo (GOJC) e a todos os participantes. O GOJC é a segunda casa de quaisquer cristãos católicos. Sabe o quanto é importante estarmos ao lado de pessoas que tem os mesmos sonhos na vida, A vida ETERNA em CRISTO! Meu agradecimento por tudo. A paz de Jesus e o Amor de Maria.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa concebida e pelo apoio financiamento. Minha eterna gratidão!

A todos amigos e colegas deixo aqui meus agradecimentos.

*Minha vida é andar por este país
Pra ver se um dia descanso feliz
Guardando as recordações
Das terras onde passei
Andando pelos sertões
E dos amigos que lá deixei*

(A Vida do Viajante - Luiz Gonzaga)

SUMÁRIO

	Páginas
1. RESUMO	1
2. SUMMARY	3
3. INTRODUÇÃO: Problemática e Objetivos	5
4. AVALIAÇÃO DE MODELOS DE RNA E SVM NA ESTIMATIVA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL (H_b) COM DIFERENTES VARIÁVEIS DE ENTRADA.	10
4.1 RESUMO.....	10
4.2 INTRODUÇÃO	11
4.3 DESCRIÇÃO DOS MÉTODOS UTILIZADOS.....	13
4.3.1 Rede Neural Artificial	14
4.3.2 Máquina de Vetor de Suporte (SVM)	15
4.3.3 Software utilizado	16
4.3.4 Cálculo das variáveis (H_0 , N , H_{sc}^d , m_r)	17
4.3.5 Índices estatísticos utilizados para a determinação da eficiência dos métodos	18
4.3.6 Descrição do local de estudo e da base de dados	20
4.4 RESULTADOS E DISCUSSÃO.....	21
4.4.1 Análise do desempenho dos modelos horário	22
4.4.2 Análise do desempenho dos modelos diário	26
4.5 CONCLUSÕES	30
4.6 REFERÊNCIAS BIBLIOGRÁFICAS	31
5. ESTIMATIVA DA FRAÇÃO TRANSMITIDA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL UTILIZANDO SVM COM ALGORITMO SEQUENTIAL MINIMAL OPTIMIZATION.	37
5.1 RESUMO.....	37
5.2 INTRODUÇÃO	38
5.3 MATERIAL E MÉTODOS	40
5.3.1 Local de estudo e dados	40
5.3.2 Máquina de Vetor de Suporte (SVM)	42
5.3.3 Modelos Estatísticos.....	43
5.3.4 Geração dos modelos estatísticos.....	44

5.3.5 Base de dados para o treinamento e validação dos modelos.....	45
5.3.6 Cálculo H_0 , do fotoperíodo (N) e H_{sc}^d	46
5.3.7 Indicadores de validação	46
5.4 RESULTADOS E DISCUSSÃO.....	48
5.4.1 Análise de precisão dos modelos SMO1, SMO2 e SMO3.....	48
5.4.2 Análise de precisão dos modelos	51
5.4.3 Análise comparativa entre a precisão de SMO e os modelos estatísticos.....	54
5.4.4 Análise de precisão de SMO nas estações chuvosa e seca: estudo de caso. ..	55
5.5 CONCLUSÕES	56
5.6 REFERÊNCIAS BIBLIOGRÁFICAS	57
6. ESTIMATIVA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL: REDES NEURAIS ARTIFICIAIS E MODELOS ESTATÍSTICOS.	63
6.1 RESUMO.....	63
6.2 INTRODUÇÃO	64
6.3 MATERIAL E MÉTODOS	65
6.3.1 Local de estudo e dados	65
6.3.2 Modelos Estatísticos.....	67
6.3.3 Redes Neurais Artificiais (RNAs).....	68
6.3.4 Software e metodologia utilizada.....	70
6.3.5 Cálculo da H_0 , do fotoperíodo e H_{sc}^d	70
6.3.6 Base de dados para validação	71
6.3.7 Avaliação estatística.....	72
6.4 RESULTADOS E DISCUSSÃO.....	73
6.4.1 Estimativas com modelos baseados na transmissividade atmosférica (kt)	73
6.4.2 Estimativas dos modelos baseados na razão de insolação	76
6.4.3 Comparação das redes neurais com os modelos estatísticos.....	77
6.4.4 Estimativa de H_b com modelos de redes neurais: Estudo de caso para condição de céu sem nuvens	78
6.5 CONCLUSÕES	80
6.6 REFERÊNCIAS BIBLIOGRÁFICAS	80
7. VARIABILIDADE TEMPORAL DA TURBIDEZ ATMOSFÉRICA E ATENUAÇÃO DE DNI EM REGIÃO DE CANA DE AÇÚCAR - BOTUCATU/SP.	85
7.1 RESUMO.....	85
7.2 INTRODUÇÃO	86

7.3 MATERIAL E MÉTODOS	87
7.3.1 Descrição do local e clima	87
7.3.2 BASE DE DADOS E MODELO	89
7.3.2.1 Descrição das medidas	89
7.3.2.2 Turbidez de Linke (TL).....	90
7.4 RESULTADOS E DISCUSSÃO.....	92
7.4.1 Variação diurna e sazonal da turbidez de Linke	92
7.4.2 Análise da turbidez de Linke média mensal.....	95
7.4.3 Comparação de TL calculada em Botucatu com as disponíveis na Website SoDA.....	99
7.4.4 Correlações polinomiais para turbidez de Linke.....	100
7.4.5 Relação entre a turbidez de Linke e os parâmetros meteorológicos	101
7.4.6 Distribuição de frequência da turbidez de Linke	102
7.5 CONCLUSÕES	104
7.6 REFERÊNCIAS BIBLIOGRÁFICAS	105
8. CONCLUSÕES GERAIS	111
9. REFERÊNCIAS BIBLIOGRAFICAS	113

LISTA DE TABELAS

CAPÍTULO I	Páginas
Tabela 1. Variáveis de entrada para os modelos de RNA e SVM.....	17
Tabela 2. Coeficientes de determinação (R^2) dos modelos de RNA e SVM.....	21
CAPÍTULO II	
Tabela 1. Modelos de SMO.....	43
Tabela 2. Modelos estatísticos gerados.....	45
Tabela 3. Base que constituem o ano típico (AT) e ano atípico (AAT).....	46
Tabela 4. Coeficiente de determinação dos modelos SMO.....	48
CAPÍTULO III	
Tabela 1. Séries climáticas da temperatura (T_{ar}), umidade do ar (UR) e precipitação (P) no período de 1971 a 2009.....	67
Tabela 2. Modelos estatísticos utilizados na comparação.....	68
Tabela 3. Índices estatísticos de validação dos modelos.....	73
CAPÍTULO IV	
Tabela 1. Tipos de atmosfera para os diferentes níveis de turbidez atmosférica (LECKER, 1978).....	92
Tabela 2. Variabilidade sazonal da turbidez de Linke em Botucatu para os anos de 1996-2008.....	95
Tabela 3. Turbidez atmosférica de Linke média mensal com TL_{Dj} e TL_{Li} para Botucatu, 1996-2008.....	98
Tabela 4. Turbidez de Linke disponível em SoDA (www.soda-is.com).....	99

LISTA DE FIGURAS

CAPÍTULO I	Páginas
Figura 1. Índices de validação dos modelos RNA e SVM na estimativa da H_b^h . a) rMBE, b) rRMSE e c) “d”	23
Figura 2. Box plots de erros dos quatorze modelos horário. a) HRNA e b) HSVM.....	25
Figura 3. Dispersão dos dados medidos e estimados por RNA e SVM para H_b^h : a) HRNA1, b) HSVM1, c) HRNA6 e d) HSVM3.....	26
Figura 4. Índices de validação dos modelos RNA e SVM na estimativa de H_b^d . a) rMBE, b) rRMSE e c) “d”	27
Figura 5. Box plots de erros dos dezesseis modelos diário. a) DRNA e b) DSVM.....	28
Figura 6. Dispersão dos dados medidos e estimados por RNA e SVM para H_b^d : a) DRNA1, b) DSVM1, c) DRNA4, d) DSVM5, e) DRNA8 e f) DSVM8.....	29
 CAPÍTULO II	
Figura 1. Localização da estação radiométrica.....	41
Figura 2. a) $kt_b^h \times kt^h$; b) $kt_b^d \times kt^d$; c) $kt_b^d \times n/N$	44
Figura 3. Comparação do valor da irradiação média entre - anos com desvio padrão e o valor da irradiação média do ano, para os meses: a) Fevereiro e b) Agosto.....	45
Figura 4. Dispersão dos valores de kt_b medido versus estimado através de: a) SMO1 e c) SMO2. Distribuição de frequência dos Resíduos (Δ): b) SMO1 e d) SMO2. Ano típico (AT) e Ano Atípico (AAT).....	49
Figura 5. Índices de validação dos SMO1 e SMO2. *resultados com ano típico (AT). **resultados com ano atípico (AAT).....	50
Figura 6. Dispersão e índices de validação do modelo SMO3 na estimativa de kt_b^d (a) e Δ (b).....	51
Figura 7. Dispersão de kt_b medido e estimado, com a frequência de Δ , respectivamente, para o modelo #M1 (a) e (b) frequência de Δ , modelo #M2 (c) e (d) frequência de Δ ...52	52
Figura 8. Índices de validação dos modelos #M1 e #M2. (*) Ano Típico e (**) Ano Atípico.....	53
Figura 9. Dispersão e valores dos índices de validação do modelo #M3 na estimativa de kt_b^d , (a) dispersão e (b) Δ	54
Figura 10. Erro percentual relativo (e, %) dos modelos: a) SMO e b) #M. (*) Ano típico e (**) ano atípico.....	55
Figura 11. Dispersão e índices de validação. a) SMO1.1, b) SMO2.2 e c) SMO3.3.....	56

CAPÍTULO III

Figura 1. Local e dados utilizados. a) localização da estação, b) Pireliômetro, c) Piranômetro e d) Heliógrafo.....	66
Figura 2. Arquitetura de uma MLP com uma camada oculta.....	69
Figura 3. Dispersão entre os valores medidos e estimados dos modelos horários a partir da relação entre $kt_b \times kt$, a) RNA1 e b) ME1. H_{bE}^h é a irradiação direta na incidência normal estimada; H_{bM}^h é a irradiação direta na incidência normal medida.....	75
Figura 4. Dispersão entre os valores medidos e estimados dos modelos diário de H_b^d obtidos a partir da relação entre $kt_b \times kt$. a) RNA2 e b) ME2. H_{bE}^h é a irradiação direta na incidência normal estimada; H_{bM}^h é a irradiação direta na incidência normal medida.....	75
Figura 5. Índices estatísticos de validação dos modelos diários de H_b^d obtidos a partir da relação entre $kt_b \times n/N$. a) rMBE, b) rRMSE e c) “d” de Willmontt.....	76
Figura 6. Dispersão entre os valores medidos e estimados dos modelos diários a partir da relação entre $kt_b \times n/N$, a) RNA3 e b) ME3. H_{bE}^h é a irradiação direta na incidência normal estimada; H_{bM}^h é a irradiação direta na incidência normal medida.....	77
Figura 7. Índices estatísticos de validação de RNA1, RNA2 e RNA3 para condição de céu claro. a) rMBE, b) rRMSE e c) “d” de Willmontt.....	78
Figura 8. Dispersão entre os valores medidos e estimados para condição de céu claro. a) RNA1, b) RNA2 e c) RNA3.....	79

CAPÍTULO IV

Figura 1. Mapa mostrando as principais áreas de produção de cana no Brasil e municípios no Estado de São Paulo. Adaptado da UNICA http://www.unica.com.br/production-map/	88
Figura 2. a) Variação horária média da turbidez de Linke e b) desvio relativo para Botucatu entre 1996-2008.....	93
Figura 3. Variabilidade sazonal da turbidez de Linke em Botucatu para os anos de 1996-2008. a) metodologia TL_{Dj} e b) metodologia TL_{Li}	94
Figura 4. Variação da turbidez de Linke em Botucatu, nos anos de 1996-2008. a) Variação Média anual. b) Variação média anual por estação.....	95
Figura 5. Médias mensais do período 2000 a 2008, considerando apenas dias de céu limpo. a) vapor d’água e b) AOD.....	97

Figura 6. Regressão polinomial da curva da variação da turbidez de Linke em Botucatu.....	100
Figura 7. Variação de TL_{Dj} e TL_{Li} em função: (a) Velocidade do vento, (b) Temperatura do ar.....	102
Figura 8. Turbidez atmosférica TL_{Dj} e TL_{Li} entre os anos 1996 e 2008. (a) Distribuição de frequência horária e (b) Frequência acumulada horária.....	103

LISTA DE SÍMBOLOS

(AMs)	Aprendizado de Máquinas.
(RNA's)	Redes Neurais Artificiais.
(AOD)	Profundidade óptica dos aerossóis (adimensional).
(CosZ)	Cosseno do ângulo zenital.
(CSP)	Sistemas de energia solar concentrada.
(CST)	Tecnologias de concentradores solares.
(d)	Índice de Willmott (adimensional).
(DNI)	Irradiância solar direta na incidência normal (Wm^{-2}).
(D _{TL})	Desvio relativo da turbidez de Linke (%).
(e)	Erro percentual (%).
(E ₀)	Fator de correção de excentricidade da órbita terrestre (adimensional).
(h)	Ângulo elevação solar (graus).
(H ₀)	Irradiação solar no topo da atmosfera (MJm^{-2}).
(H _b)	Irradiação direta na incidência normal (MJm^{-2}).
(H _d)	Irradiação solar difusa (MJm^{-2}).
(H _g)	Irradiação solar global (MJm^{-2}).
(H _{sc} ^d)	Irradiação solar direta no topo da atmosfera (MJm^{-2}) (no dia).
(I _G)	Irradiância solar global ($\text{Js}^{-1} \text{m}^{-2} = \text{W m}^{-2}$).
(I _{sc})	Constante solar integrada no topo da atmosfera ($4,92 \text{ MJm}^{-2}$) (na hora).
(kt)	Transmissividade atmosférica (adimensional).
(kt _b)	Fração transmitida da irradiação direta na incidência normal (adimensional).
(m _a)	Massa ótica a pressão real (adimensional).
(MBE)	Mean Bias Error.
(MLP)	Multilayer Perceptron.
(MODIS)	Moderate Resolution Imaging Spectroradiometer.
(m _r)	Massa ótica relativa do ar (adimensional).
(N)	Fotoperíodo (horas).
(n)	Insolação solar (horas).
(r'=n/N)	Razão de insolação (adimensional).
(RBF)	Radial Basis Function.
(R _D)	Irradiância solar difusa (Wm^{-2}).
(I _G)	Irradiância solar global (Wm^{-2}).

(rMBE)	Relative Mean Bias Error (%).
(RMSE)	Root Mean Square Error.
(rRMSE)	Relative Root Mean Square Error (%).
(SMO _{reg})	Sequential Minimal Optimization.
(SVM's)	Support Vector Machines ou Máquinas de Vetores de Suporte.
(SVR)	Support Vector Regression.
(TL)	Fator de Turbidez de Linke (adimensional).
(TL _{Dj})	Turbidez de Linke (metodologia de Djafer e Irbah).
(T _{lk})	Fator de correção da turbidez de Linke (adimensional).
(TL _{Li})	Turbidez Linke (metodologia de Li e Lam) (adimensional).
(WEKA)	Waikato Environment for Knowledge Analysis.
(ZCAS)	Zona de Convergência do Atlântico Sul.
(C , γ e ϵ)	Parâmetros da RBF.
[DNI(h)]	Irradiância solar direta na incidência medida (Wm^{-2}) (na hora).
[I_0 (h)]	Constante solar ($=1367 \text{ Wm}^{-2}$) (na hora).
[δ_{Ra} (m_a)]	Espessura óptica integrante (adimensional).
[δ_{Rk} (m_a)]	Espessura óptica integrante de Rayleigh (adimensional).
R_0 e R	São a distância instantânea e a distância média entre a terra-sol.
(E_0)	Fator de correção de excentricidade da órbita terrestre (adimensional).

MODELAGEM DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL EM BOTUCATU: APRENDIZADO DE MÁQUINA, ESTATÍSTICO E LINKE. Botucatu, 2016. 127p. Tese (Doutorado em Agronomia/Irrigação e Drenagem) – Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista.

Autor: CÍCERO MANOEL DOS SANTOS

Orientador: Dr. JOÃO FRANCISCO ESCOBEDO

1. RESUMO

A irradiação direta na incidência normal (H_b) possui papel importante no manejo de culturas agrícolas, na utilização como fonte de energia renovável e na modelagem atmosférica. Apesar de sua importância em diferentes áreas, medidas pontuais de H_b não são facilmente disponíveis nos centros de pesquisas, devido ao elevado custo de exportação dos sensores e suas manutenções periódicas. Os modelos estatísticos têm sido desenvolvidos e utilizados para estimativa de H_b nos locais onde não são monitorados. Estes modelos, normalmente, utilizam a H_g como variável de entrada, pois é a variável mais comumente medida em estações solarimétricas. Os modelos estatísticos correlacionam à fração transmitida da irradiação direta na incidência normal (kt_b) com transmissividade atmosférica (kt) ou com a razão de insolação (n/N). Recentemente as técnicas de Aprendizado de Máquinas foram inseridas para estimativa de H_b . Teoricamente, são técnicas que apresentam alto desempenho na estimativa de modelos e gerar valores estimados mais precisos de H_b que os modelos estatísticos. O trabalho está dividido em 4 capítulos divididos da seguinte forma. Capítulo 1: Propor a utilização da técnica Máquina de Vetor de Suporte – SVM e da Redes Neurais Artificiais para estimativa de H_b e comparar com os modelos estatísticos, testando diferentes variáveis de entrada, . Capítulo 2: Comparar a SVM com os modelos estatísticos. Capítulo 3: Comparar Rede Neural Artificial – RNA com os modelos estatísticos, utilizando o algoritmo *Backpropagation*. Capítulo 4: Modelagem da turbidez atmosférica de Linke com H_b . A fração transmitida de H_b (kt_b) é modelada para obter H_b . Para treinamento e validação dos modelos é utilizado um banco de dados de 13 anos (1996 – 2008), medidos na estação radiométrica localizada na Faculdade de Ciências Agrônômicas – FCA/UNESP (22,85°S; 48,45°W e 786m). Foram testadas diferentes variáveis de entrada para verificar qual a melhor na estimava

dos modelos. Os índices estatísticos: *MBE*, *rMBE*, *RMSE*, *rRMSE*, *d* de Willmott e o erro percentual (%) são utilizados para validar os modelos. Os modelos foram propostos e avaliados nas partições de tempo: horária e diária. Os resultados mostraram que os modelos estatísticos estimam H_b com resultados ($20\% \leq rRMSE < 30\%$). Os modelos propostos (SVM e RNA) geram resultados melhores que os modelos estatísticos e são indicados para estimativa de H_b ($rRMSE < 20\%$). O modelo da SVM estima H_b melhor que RNA, por isso seu uso é tido como a primeira escolha entre os modelos.

Palavras – chave: Redes Neurais Artificiais, Máquina de Vetores de Suporte, Variáveis Climáticas.

MODELING OF DIRECT IRRADIATION AT NORMAL INCIDENCE IN BOTUCATU: MACHINE LEARNING, STATISTICAL AND LINKE. Botucatu, 2016. 127p. Thesis (Ph. D. in Agronomy/Irrigation and Drainage) – Faculty of Agricultural Sciences, University of São Paulo State.

Author: CÍCERO MANOEL DOS SANTOS

Advisor: Dr. JOÃO FRANCISCO ESCOBEDO

2. SUMMARY

The direct irradiation at normal incidence (H_b) is an important role in the management of crops, in the use as a renewable energy source and atmospheric modeling. Despite its importance in different areas, specific measures H_b are not readily available in research centers, due to the high cost of exporting the sensors and periodic maintenance of the sensors. Statistical models have been developed and used to estimate H_b in places where they are not monitored. These models usually use the H_g as input variable, as is the variable most commonly measured in solarimetric stations. Statistical models correlate to the fraction transmitted at H_b (kt_b) with atmospheric transmissivity (kt) or insolation ratio (n/N). Recently the Machine Learning techniques (ML) were inserted for estimation of H_b . Theoretically, these techniques have greater capacity to model and generate more precise values of H_b than statistical models. The work is divided into four chapters divided as follows. Chapter 1: To propose the use of Support Vector Machine (SVM) and Artificial Neural Networks (ANN) technique to estimate H_b and compare the statistical models, testing different input variables. Chapter 2: To compare the SVM with the statistical models. Chapter 3: To compare Artificial Neural Network (ANN) with statistical models using the backpropagation algorithm. Chapter 4: Modeling of atmospheric turbidity Linke with H_b . The kt_b is modeled to get indirectly H_b . The validation methodology of the models with typical and atypical year is adopted and evaluated. It used a database of 13 years data (1996-2008), measured in radiometric station located at the Faculty of Agricultural Sciences - FCA/UNESP (22.85° S, 48.45° W and 786m. Different input variables are tested in the models to see if the estimate is improving. The variables used are: H_b , H_g , solar insolation (n), air temperature and relative humidity the other variables were obtained by mathematical equations. Statistical indices: MBE , $rMBE$, $RMSE$, $rRMSE$, d

Willmontt and percent error (%) are used to validate the models. The models are proposed and evaluated in time: hourly and daily partitions. The results show that the statistical models estimate H_b with acceptable results ($rRMSE \leq 20\%$ $<30\%$). The proposed models (SVM and ANN) generate better results than the statistical models and are suitable for estimation of H_b ($rRMSE <20\%$). The model of SVM estimates H_b better than ANN, so its use is considered the first choice among the models.

Keywords: Neural Artificial Network, Support Vector Machine, Climatological Variables.

3. INTRODUÇÃO: Problemática e Objetivos

A sociedade contemporânea caracteriza-se, sobretudo, pelo elevado desenvolvimento da ciência e tecnologia e, conseqüentemente, do aumento populacional. Neste sentido, vive-se em um mundo em que a robótica, a informática e o consumo energético são requisitos que ditam e expressam o quão desenvolvido é um país.

O Brasil, inserido neste contexto, e diante da globalização e do capitalismo sem fronteiras, já usufrui de tais inovações, embora, na maioria dos casos, seja mero importador. Devido ao crescente consumo energético e as inúmeras estiagens na última década, o Brasil sofreu com numerosos apagões devido à matriz energética brasileira, praticamente, ser de origem hídrica.

Na expectativa de suprir a demanda energética nacional diferentes pesquisas para o aproveitamento de fontes alternativas de energia promissoras/renováveis ganharam destaque, tais como eólica, energia das marés, geotérmica, biomassa e, principalmente, a energia de origem solar.

Com uma distância média de ($D \approx 1,5 \times 10^8$ km) entre a Terra e o Sol, praticamente toda a energia que chega à Terra (topo da atmosfera e superfície terrestre) é que define o clima, as atividades humanas e os processos que ocorrem no sistema solo-planta-atmosfera (LIOU, 2002). O Sol é uma esfera com raio de ($R \approx 6,96 \times 10^5$ km) e massa ($M \approx 1,99 \times 10^{30}$ kg), toda radiação solar é gerada pela conversão

de quatro átomos de hidrogênio estáveis para um átomo de hélio em reações termo-nucleares.

A radiação solar [que inclui a luz visível ($0,40 \leq \lambda < 0,76 \mu\text{m}$), infravermelho ($0,76 \leq \lambda \leq 4,00 \mu\text{m}$) e a radiação ultravioleta ($0,10 \leq \lambda < 0,40 \mu\text{m}$)] é transmitida para o espaço em todas as direções e somente uma pequena fração chega ao topo da atmosfera terrestre (IQBAL, 1983). Ao interagir com a atmosfera, cerca de 30% da radiação é absorvida ou espalhada pelos seus constituintes. A absorção ocorre, em grande parte, na faixa do ultravioleta, pelo ozônio, pelo vapor d'água, dióxido de carbono, pelas moléculas e os átomos em comprimento de onda discreto (LOUCHE et al., 1988). A absorção por ozônio ocorre na estratosfera, acima das camadas onde outras moléculas gasosas ou processos de espalhamento por aerossóis ocorrem (GUEYMARD, 2003). O espalhamento é causado pela interação da radiação com aerossóis, principalmente. As nuvens também contribuem com a reflexão e absorção, principalmente, quando combinam o tipo, número e espessura da camada.

A radiação solar que chega a Terra (topo da atmosfera até a superfície terrestre) de todas as direções por unidade de área e por unidade de tempo é definida como irradiância solar global (I_G , $\text{Js}^{-1} \text{m}^{-2} = \text{W m}^{-2}$), que ao ser integrada em intervalos de tempo específicos define-se irradiação solar global (H_g , MJ m^{-2}).

A H_g é a soma da irradiação direta (H_D) com irradiação difusa (H_d). A H_d possui sua direção alterada por espalhamento na atmosfera terrestre. Convertida para um plano de incidência normal através da divisão pelo cosseno do ângulo zenital, a H_D torna-se irradiação direta na incidência normal (H_b). Regida pela lei de Beer – Bouguer – Lambert, a H_b é emitida para a superfície da terra com o mesmo ângulo que foi incidente na atmosfera. De acordo com a lei de Beer – Bouguer – Lambert, a atenuação da luz através de um meio é proporcional à distância percorrida no meio e ao fluxo local de radiação. Neste sentido, a H_b é a energia total incidente por unidade de tempo, numa unidade de área perpendicular à superfície terrestre na direção do disco solar (HALTHORE et al., 1997).

Conhecer as regiões que possuem maior incidência de H_b é extremamente importante em estudos de energias renováveis, tomadas de decisões no manejo agrícola e estudos climáticos. No Brasil, a conversão da energia solar em energia elétrica é uma alternativa de energia renovável para a complementação da matriz energética, devido à grande extensão territorial ($A \approx 8.515.767,049 \text{ km}^2$) e localização geográfica.

A H_b é a principal componente no processo de conversão fotovoltaico (conversão direta da energia solar em elétrica), nas aplicações térmicas (aquecimento de água e iluminação natural de ambientes) para residências e centros comerciais. Numa perspectiva futura, espera-se que mais uma porcentagem maior da H_b que chega a superfície terrestre seja aproveitada nas novas tecnologias energéticas.

Na agricultura, a H_b é uma das variáveis responsáveis pelo processo de Evapotranspiração (ET) e está relacionado com a produtividade potencial da cultura. Pequena fração da H_b incidente é utilizada para a realização da fotossíntese, parte é absorvida pela clorofila e outra fração para o crescimento das plantas (MONTEITH, 1977; ALLEN, 1996; PEREIRA et al., 2002). H_b é parâmetro fundamental, também, no balanço hídrico e manejo adequado de irrigação. Aplicado no contexto agrônomo, a H_b assume grande importância quando é utilizada em painéis fotovoltaicos em ambientes automatizados controlando sistemas de irrigação.

Uma longa série de H_b pode ser obtida através de um pireliômetro de termopilha acoplado a um rastreador solar instalado em estações solarimétricas de superfície (GUEYMARD, 2010), com o monitoramento contínuo e manutenção periódica. Indiretamente, H_b pode ser obtida da diferença entre medidas de H_g e da irradiação solar difusa (H_D). Contudo, as componentes H_D e a H_b são poucas monitoradas nos principais centros de pesquisa do mundo. A escassez de medidas de H_b e H_D são ocasionadas pelo elevado custo de aquisição e manutenção dos sensores e seus itens adjacentes (cabos conectadores). Outra dificuldade na obtenção de medidas de H_D , consiste na correção mediante um fator multiplicativo que leva em conta a largura e o raio do anel de sombreamento, a latitude do local, a declinação solar, o ângulo do nascer do sol e um fator para a condição de nebulosidade (OLIVEIRA et al., 2001), que eleva o erro da medida.

Embora seja essencial nos projetos e na simulação de sistemas de energia solar, em aplicações para concentradores para elevar a temperatura do sistema (energia térmica), em sistemas fotovoltaicos, no monitoramento de tendências de turbidez atmosféricas e na poluição durante longos períodos (Batlles et al., 2000; Gueymard, 2003), a H_b é pouco monitorada quando comparada com a irradiação solar global (H_g), a insolação solar (n), temperatura do ar e umidade relativa.

Na ausência de sensores, a maneira mais prática de se obter valores confiáveis de H_b é através da modelagem. Diferentes modelos têm sido propostos e utilizados na estimativa de H_b : Os modelos de natureza física ou de

transferência radiativa (BIRD:HULSTROM, 1980; YANG et al., 2001; GUEYMARD, 2012; BEHAR et al, 2015) e modelos estatísticos ou empíricos que relacionam a H_b com a transmissividade atmosférica ($k_t = H_g/H_0$) [H_0 é a irradiação solar no topo da atmosfera) e a razão insolação solar (n/N) [N é o fotoperíodo] (ORGILL et al., 1977; REINDL et al., 1990; LOUCHE et al., 1991; BATLLES et al., 2000; INEICHEN, 2008). Recentemente, com o desenvolvimento das ferramentas computacionais e de novos métodos estatísticos, as técnicas de Aprendizado de Máquinas (AM) ganharam grande destaque na modelagem solar.

As técnicas AM têm sido amplamente utilizadas e vêm mostrando ser uma ferramenta poderosa na previsão e regressão de dados em diversas áreas. Entre as técnicas de AM a Rede Neural Artificial (RNA) e a Máquina de Vetores de Suporte (SVM) são as mais utilizadas. A RNA, que está cada vez mais recebendo atenção na resolução de problemas complexos, é conhecida como funções aproximadores universais e são capazes de aproximar todas as funções não-lineares contínuas para precisão arbitrária (Jian, 2009). Por sua natureza robusta e sua capacidade de compensar erros sistemáticos, nos últimos anos a Multilayer Perceptron (MLP), um algoritmo de treinamento de RNA, é usado por muitos pesquisadores para estimar a H_g (Yadav e Chandel, 2014).

A SVM, originalmente, foi desenvolvida por Vapnik (1995) e tem sido amplamente aplicada em diferentes áreas, tais: engenharia, informática, climática (Raghavendra e Deka, 2014; Luna et al., 2014; Mohammadi et al., 2015). As técnicas de AM já estão implementadas em softwares gratuitos e de código aberto para reduzir o tempo dos usuários.

Apesar do sucesso em muitas áreas, não existem na literatura trabalhos sobre a aplicação da SVM na estimativa da H_b e poucos são os trabalhos utilizando RNA (Alam et al., 2006; Tomar et al., 2012; Kaushina et al., 2014). Para o Brasil escasso são os trabalhos utilizando RNA e SVM na estimativa da H_b . Estimativas precisas de H_b são de extrema importância para a verificação das regiões com grande potencial solar e para instalações de futuras de usinas solares. Avaliando resultados numa microescala é um passo muito importante para aplicação e avaliação dos modelos numa escala maior.

Diante das considerações, esta pesquisa tem como objetivos: 1- gerar e avaliar modelos de estimativa de H_b com base nas relações estatísticas que correlacionam a fração transmitida da irradiação direta na incidência normal (k_{tb}) com

kt (Liu e Jordan, 1960) e $kt_b \times n/N$ (Ångström, 1924), nas partições horária e diária; 2- avaliar a potencialidade e propor a utilização de SVM na estimativa de H_b , a partir de kt_b ; 3- Avaliar e propor a utilização de RNA na estimativa de H_b , a partir de kt_b ; 4- Comparar os resultados dos modelos estatísticos com os modelos gerados com SVM e RNA. 5 – Avaliar diferentes variáveis de entrada na estimativa de H_b . 5- Indicar a melhor metodologia para estimativa de H_b . 6- Testar uma nova metodologia de validação, ano típico e atípico, selecionado no banco de dados; 7- Avaliar o efeito dos aerossóis e vapor d'água em H_b , a partir do modelo paramétrico.

O estudo será desenvolvido em Botucatu – SP. Para atingir estes objetivos a tese está dividida em 4 capítulos, o primeiro capítulo intitulado **“AVALIAÇÃO DE MODELOS DE RNA E SVM NA ESTIMATIVA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL (H_b) COM DIFERENTES VARIÁVEIS DE ENTRADA”**; o segundo capítulo intitulado: **“ESTIMATIVA DA FRAÇÃO TRANSMITIDA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL UTILIZANDO SVM COM ALGORITMO SEQUENTIAL MINIMAL OPTIMIZATION”**; o terceiro capítulo com título **ESTIMATIVA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL: REDES NEURAIAS ARTIFICIAIS E MODELOS ESTATÍSTICOS**; o quarto capítulo com título **“VARIABILIDADE TEMPORAL DA TURBIDEZ ATMOSFÉRICA E ATENUAÇÃO DE DNI EM BOTUCATU/SP, REGIÃO COM QUEIMADAS”**.

4. AVALIAÇÃO DE MODELOS DE RNA E SVM NA ESTIMATIVA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL (H_b) COM DIFERENTES VARIÁVEIS DE ENTRADA.

4.1 RESUMO

Dados de irradiação direta na incidência normal (H_b) são essenciais para calibração de satélites, projetos e simulações de plantas de usinas solares concentradas. As medidas de H_b são escassas, por isso, estimativas com modelos são recomendadas. Este trabalho estima H_b utilizando Rede Neural Artificial (RNA) e Máquina de Vetores de Suporte (SVM). São avaliados quinze modelos gerados de RNA e Quinze modelos gerados de SVM com diferentes variáveis de entrada. As diferentes variáveis de entrada também são avaliadas. São utilizadas RNA do tipo Perceptron de Múltiplas camadas (MLP). A função *Radial Basis Function* (RBF) de Kernel é adotada na SVM. Os modelos foram avaliados nas escalas horária e diária. Dados de 13 anos, medidos numa estação solarimétrica localizada em Botucatu/SP, foram utilizados para o treinamento e validação. O *Relative Mean Bias Error* (rMBE), *Relative Root Mean Square Error* (rRMSE), o coeficiente de determinação (R^2) e d de Willmott foram usados para na avaliação dos modelos. Os valores de R^2 obtidos dos modelos são elevados. Dentre as variáveis de entrada avaliadas, a transmissividade atmosférica da irradiação global (kt), a constante solar integrada (H_{sc}) e a razão de insolação (n/N) foram as mais relevantes nos modelos. Os valores de rMBE e rRMSE, nas duas partições de tempo, dos modelos de SVM são menores que os valores obtidos com a RNA. Os modelos horários possuem

maiores valores de rRMSE que os modelos diários. Os modelos melhoraram as estimativas de H_b em relação a outros resultados da literatura. A SVM possui melhor desempenho na estimativa de H_b que a RNA. Portanto, as estimativas são mais precisas e a SVM é recomendada como a primeira opção na estimativa de H_b com bons valores.

Palavras – chave: Radiação solar; aprendizado de máquina; WEKA.

4.2 INTRODUÇÃO

Estudos recentes mostram a relevância de medidas confiáveis de irradiação direta na incidência normal (H_b) para calibração de satélites (AL-RASHEEDI et al., 2014), para modelagem atmosférica e sob condição de céu sem nuvens na conversão em energia elétrica a partir das tecnologias de concentradores solares (CST), como os sistemas de energia solar concentrado (CSP) e os sistemas fotovoltaicos concentrados (CSPV) (LAW et al., 2014). O interesse de H_b para tecnologia CSP é devido à irradiação difusa (H_D) não ser eficientemente concentrada, pois os sistemas de energia solar são mais eficientes em condições de céu sem nuvens do que em condições de céu com nebulosidade (GUEYMARD, 2003). Na modelagem atmosférica a H_b é destaque em estudos climáticos e de poluição, com aplicação em conforto térmico de edifícios e iluminação natural de ambientes. Devido ao preço competitivo no mercado de energia e a redução de custo, nos últimos anos, a energia solar concentrada e fotovoltaica têm se tornado uma das principais fontes energéticas alternativa para futuro. Portanto, medidas de longo prazo de H_b são essenciais para o planejamento financeiro, a análise de desempenho e desenvolvimento das tecnologias de sistemas de energia solar.

Embora existam conjuntos de dados de radiação solar e vários mapas solarimétricos no mundo, estes geralmente não são detalhados o suficiente para ser usado para a determinação de energia solar disponível em áreas pequenas (BADESCU et al., 2013). Além disso, mapas solares e medidas de H_b não são facilmente disponíveis na maioria dos locais do mundo. As medidas de H_b são obtidas através de pireliômetros. O grande problema encontrado para obtenção de medidas de H_b reside no elevado custo de aquisição do sensor, dos sistemas de rastreamentos, dos dispositivos adjacentes e da necessidade de manutenção periódica para garantir a

precisão das medidas (BERTRAND et al., 2015). Na ausência destes sensores, estudos têm sido realizados para recuperação de série histórica e fornecer estimativas nos locais onde as medições não são realizadas ou não estão facilmente disponíveis.

A estimativa, normalmente, é a partir de modelos baseados na transferência radiativa (códigos computacionais sofisticados) ou utilizando modelos de decomposição. Os modelos baseados nos complexos princípios físicos de transferência radiativa são úteis na condição de céu sem nuvens e levam em consideração a atenuação de H_b por aerossóis, espalhamento de Rayleigh, absorção pelo ozônio, absorção pelo vapor d'água e gases (GUEYMARD, 2003). Porém, o uso dos modelos de transferência radiativa num local de interesse é limitado devido à escassez dos dados de entrada necessários para a modelagem. Já os modelos de decomposição são empíricos por natureza, ajustados localmente e calculam H_b a partir da irradiação global (H_G). Estes modelos correlacionam as frações das componentes de H_b (fração transmitida, k_t), de H_G (transmissividade atmosférica, k_t), de H_D (fração difusa, k_D) ou da razão de insolação ($r' = n/N$) (GUEYMARD:RUIZ-ARIAS, 2014). Quando são utilizadas muitas variáveis de entrada, os modelos de composição tornam-se complicados, demorados e o uso de regressão linear múltiplo torna-se inadequado (LOPEZ et al., 2005). No entanto, estes modelos têm incertezas intrínsecas, que tornam-se limitada seu uso. Portanto, a estimativa de H_b não é tão simples. Assim, novas metodologias são necessárias quando as existentes tornam-se limitadas ou ineficientes para algumas situações.

Por isso, técnicas de aprendizado de máquinas (AM) têm sido utilizadas na modelagem solar e vêm mostrando ser boa ferramenta na previsão e regressão de dados. Existem várias técnicas de AM, incluindo as Rede Neural Artificial (RNA) e Máquina de Vetor de Suporte (SVM), que são as mais utilizadas (JIAN, 2009; YADAV:CHANDEL, 2014; RAGHAVENDRA:DEKA, 2014; LUNA et al., 2014; MOHAMMADI et al., 2015b). As técnicas de AM já estão disponibilizadas em softwares gratuitos e de código aberto para facilitar a utilização e o tempo dos usuários. O software *Waikato Environment for Knowledge Analysis* (WEKA), por exemplo, possui seu conjunto de algoritmos de AM amplamente utilizado com sucesso em diversas áreas (TONG:ZHU, 2009; TÜFEKCI, 2014; YADAV:CHANDEL, 2015). Por não necessitarem de suposições prévias sobre as relações entre os dados de entrada, as técnicas RNA e SVM estão implementadas no WEKA e podem ser facilmente utilizados.

Apesar do sucesso de aplicação em muitas áreas, escassos são os trabalhos referentes à aplicação da SVM na estimativa de H_b e poucos são os trabalhos utilizando RNA (ALAM et al., 2006; TOMAR et al., 2012; KAUSHINA et al., 2014). No Brasil, inclusive, trabalhos utilizando RNA e SVM na estimativa da H_b são escassos ou inexistentes. Portanto, as aplicações de RNA e, principalmente, SVM na área das energias renováveis têm sido mínimo (RAMLI et al., 2015). Assim, os principais objetivos deste estudo são: - Analisar a estabilidade, a precisão e explorar o potencial de RNA e SVM na estimativa de H_b ; - Investigar as principais variáveis de entrada na modelagem de H_b ; - Comparar e indicar a melhor técnica para modelagem de H_b . Para uma análise mais ampla, os modelos são avaliados nas partições horárias e diárias. Os resultados servem como um estudo de caso, pois o governo brasileiro tem crescente interesse na utilização de novas energias renováveis para suprir a matriz energética atual (hidroelétrica). O Brasil é o quinto país mais extenso do mundo, porém, apenas alguns centros de pesquisas medem H_b de rotina. Assim, tem-se a necessidade de novas metodologias para o mapeamento de H_b num local e expandir para todo território. O trabalho está dividido em seções. A seção 2 mostra brevemente às técnicas RNA e SVM utilizadas, o software utilizado, as variáveis de entrada e os índices de validação. O local e as características do conjunto dos dados na modelagem são apresentados na seção 3. Na seção 4 são discutidos os resultados do trabalho. O trabalho é concluído na seção 5.

4.3 DESCRIÇÃO DOS MÉTODOS UTILIZADOS

A metodologia é dividida em duas partes. A primeira consiste na estimativa da irradiação direta na incidência normal usando RNA. A segunda parte envolve a estimativa de H_b utilizando SVM. O local para estudo de caso selecionado é Botucatu, cidade localizado no estado de São Paulo/Brasil. Na estimativa de H_b é necessário organizar o conjunto de dados e escolher adequadamente o algoritmo. A descrição do local e dos dados utilizados é apresentada na seção 3. Os modelos de RNA e SVM foram testados e avaliados através da ferramenta computacional WEKA (seção 2.3).

4.3.1 Rede Neural Artificial

A RNA é um sistema computacional de processamento de informações que consistem num grupo interconectado de neurônios artificiais que se baseiam na estrutura, no método de processamento e na capacidade de aprendizagem do cérebro (HAYKIN, 1998). A RNA é capaz de armazenar o conhecimento e compreender a complexa relação não-linear entre os dados de saída e os de entrada, abrangendo problemas de regressão, modelos de previsões e outras aplicações em diferentes áreas (KHALIL et al., 2005; FIORIN et al., 2011). Elas são compostas de unidades de processamento simples e, normalmente, estruturadas em três camadas distintas: entrada que apresenta os dados selecionados para o modelo, a camada oculta que processa os dados e a camada de saída que produz os resultados (LORENA et al., 2011). Por ter alta capacidade de auto-organização e processamento temporal, a RNA resolve problemas de alta complexidade (LUNA et al., 2014). Dentre as redes neurais existentes, o Perceptron de Múltiplas Camadas (MLP) é bastante utilizado.

A MLP é um sistema de processamento de informações massivamente paralelo e distribuído, aplicado com êxito a um modelo para muitos problemas não-lineares (BISHOP, M. 1995). A MLP aprende com exemplos de dados por amostragem e são conhecidas por serem aproximadores universais de uma grande classe de funções que determinam o melhor peso, e em alguns casos a estrutura da rede (PANIAGUA-TINEO et al., 2011). A estrutura básica da MLP trata-se de uma camada de entrada (x_{ij}), camada oculta com pesos de ligações (w_{ij}) e uma camada de saída (y_i). A saída da rede (y_i) pode ser modelada conforme (LAM et al., 2008):

$$y_i = \sum_{j=1}^{\eta} w_{i,j} x_{i,j} + \theta_i \quad (1)$$

onde $x_{i,j}$ é o sinal de entrada a partir de neurônios de ordem j (para a camada de entrada), $w_{i,j}$ é o peso da ligação direta do neurônio j para neurônio i (na camada oculta) e θ_i é a bias do neurônio i . A saída dos neurônios é calculada pela aplicação de uma função de ativação. A função de ativação utilizada é tipicamente sigmóide padrão (Equação 2).

$$f(x) = \frac{1}{[1+\exp(-x)]} \quad (2)$$

Existem muitos algoritmos de aprendizado, específicos para determinados modelos de redes neurais. Neste trabalho a MLP foi treinada usando o algoritmo de treinamento *backpropagation* e o termo *momentum*. Este algoritmo baseia-se na regra de aprendizagem por correção de erros e consiste em duas passagens pelas diferentes camadas das redes: um passe para frente, a propagação e um passe para trás, a retropropagação. No passo para frente, um padrão de atividade é aplicado aos nós sensoriais da rede e os seus efeitos se propagam através da rede camada por camada. O termo momento é utilizado para aumentar a taxa de aprendizado (HAYKIN, 1998; HAYKIN, 2001).

4.3.2 Máquina de Vetor de Suporte (SVM)

Máquina de Vetor de Suporte (SVM) é um método de aprendizagem de máquinas derivado da teoria de aprendizagem estatística introduzidos por Vapnik (1995), desenvolvido a partir da hipótese de minimização do risco estrutural para minimizar o risco empírico (ou de treinamento) e o intervalo de confiança, a fim de alcançar uma boa capacidade de generalização (RAGHAVENDRA:DEKA, 2014). A SVM tem ganhado destaque em muitas áreas do conhecimento para solução de complexos problemas de reconhecimento de padrões, classificação, análise de regressão e previsão (MOHAMMADI et al., 2015a).

A regressão com SVM consiste em estimar uma função de acordo com um determinado conjunto de dados definidos: $\{(x_i, y_i)\}_{i=1}^{\eta}$, onde $x_i \in \mathcal{R}^{\eta}$ é o vetor de caracterização que representa as variáveis de entrada; $y_i \in \{-1, +1\}$ é a saída e η é o número total de conjuntos de dados (TABARI et al., 2012). Neste estudo, os vetores de entrada x_i referem-se à irradiação global (H_G), a transmissividade atmosférica de H_G (kt), a constante solar integrada no topo da atmosfera (H_{sc}), a irradiação solar no topo da atmosfera H_0 , a massa ótica relativa do ar (m_r), o cosseno do ângulo zenital ($\cos Z$), a irradiação direta no topo da atmosfera diária (H_{sc}^d), temperatura do ar (máxima, mínima e média), umidade relativa do ar (UR), duração do brilho solar e a razão de insolação ($r' = n/N$) (n é duração do brilho solar e N o fotoperíodo), y_i refere aos valores de H_b . Maiores detalhes da teoria e da evolução da SVM podem ser encontrados em (VAPNIK, 1998).

O desempenho e a capacidade de aprendizado da SVM na regressão são atribuídos ao uso do conjunto de funções de Kernel, que diagramam as

informações para um espaço de dimensão mais elevado (PIRI et al., 2015), isso faz com que a SVM seja uma escolha viável para abordar vários estudos de radiação solar, que é por natureza não-linear (RAMEDANI et al., 2014). Embora novas funções estejam sendo propostas, as quatro funções básicas de Kernel são: linear, polinomial, Radial Basis Function (RBF) e sigmoidal. Pela eficiência computacional, simplicidade e adaptação para otimização de problemas mais complexos, neste estudo a função Radial Basis Function (RBF) de Kernel é considerada para a regressão de SVM.

Para permitir maior flexibilidade na aplicação de RBF é necessário ajustar adequadamente os parâmetros C (custo), γ (gamma) e ε (épsilon). Os parâmetros C e γ são mutuamente dependentes entre si, baixos valores de C produzem um aprendizado de máquina com má aproximação e valores elevados geram um aprendizado de máquina mais complexo (RAGHAVENDRA:DEKA, 2014). Quanto mais baixo o valor de γ maior o número de vetores de suporte e complexa será a aprendizagem para a máquina. Neste estudo, os valores de C e γ são testados pelo método da tentativa e aqueles com a melhor precisão na validação-cruzada são escolhidos. O parâmetro ε é aplicado para ajustar os dados de treinamento.

A SVM utilizada é a compilação LibSVM integrada para a classificação de vetores de suporte, regressão e estimativa de distribuição (disponível no site: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). Maiores detalhes de sua estrutura funcional são encontrados em (HSU et al., 2003; CHANG:LIN, 2011). A LibSVM é implementada e utilizada através do *Waikato Environment for Knowledge Analysis* (WEKA).

4.3.3 Software utilizado

O *Waikato Environment for Knowledge Analysis* (WEKA) é um conjunto de algoritmos de AMs para tarefas de mineração de dados, disponível em: <http://www.cs.waikato.ac.nz/~ml/weka/> (HALL et al., 2009). Fornecendo ao usuário uma programação em Java, o WEKA contém ferramentas para pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização. O WEKA possui a opção da escolha da porcentagem de dados para o treinamento e para a validação, definida como Percentage Split. A escolha dos dados a partir dessa ferramenta é aleatória, sem tendência na modelagem. Considerando o Percentage Split do WEKA, 70% da série foi utilizada para o treinamento e 30% para a validação e teste.

O algoritmo J48, implementado no WEKA, é utilizado para classificar as variáveis de entrada mais relevantes para estimar H_b .

As correlações entre H_b com outras variáveis de entrada são mostradas na Tabela 1. Os dados de entrada para os modelos de RNA e SVM consistem em variáveis dependentes e independentes. Diferentes combinações com as variáveis de entrada foram formadas nas partições horária e diária. No trabalho os modelos treinados com a MLP são simbolizados como RNA e com SVM os modelos da técnica SVM.

Tabela 1. Variáveis de entrada para os modelos de RNA e SVM.

		RNA	SVM	Variável de Entrada
Horário	H_b^h	HRNA1	HSVM1	H_G^x
		HRNA2	HSVM2	H_G^x, k_t^x
		HRNA3	HSVM3	H_G^x, k_t^x, H_{sc}
		HRNA4	HSVM4	$H_G^x, k_t^x, H_{sc}, H_0^x$
		HRNA5	HSVM5	$H_G^x, k_t^x, H_{sc}, H_0^x, m_r$
		HRNA6	HSVM6	$H_G^x, k_t^x, H_{sc}, H_0^x, m_r, \text{CosZ}$
		HRNA7	HSVM7	$k_t^x, H_{sc}, H_0^x, m_r, \text{CosZ}$
Diário	H_b^d	DRNA1	DSVM1	H_G^x
		DRNA2	DSVM2	H_G^x, k_t^x
		DRNA3	DSVM3	H_G^x, k_t^x, H_{sc}^d
		DRNA4	DSVM4	$H_G^x, k_t^x, H_{sc}^d, H_0^x$
		DRNA5	DSVM5	$H_G^x, k_t^x, H_{sc}^d, H_0^x, r^x$
		DRNA6	DSVM6	$k_t^x, H_{sc}^d, H_0^x, r^x$
		DRNA7	DSVM7	Tmax, Tmin, Tmed, UR, n
		DRNA8	DSVM8	$H_G^x, \text{Tmax}, \text{Tmin}, \text{Tmed}, \text{UR}, n$

H_b : é a irradiação direta na incidência normal (MJ m^{-2}); k_t : é a transmissividade atmosférica de H_G ; H_{sc} : a constante solar integrada no topo da atmosfera ($4,92 \text{ MJm}^{-2}$); H_0 : é a irradiação solar no topo da atmosfera (MJm^{-2}); m_r : é a massa ótica relativa do ar; CosZ : é o cosseno do ângulo zenital; H_{sc}^d : é a irradiação direta no topo da atmosfera (MJm^{-2}) e r^x : é a razão de insolação (n/N). x : representa a partição de tempo (horário – h, diário – d).

4.3.4 Cálculo das variáveis (H_0 , N , H_{sc}^d , m_r)

A irradiação solar no topo da atmosfera (H_0) foi calculada de acordo com equações descritas em (SOUZA et al., 2005), Eq. (3):

$$H_0 = I_0 \left[1 + 0,033 \cos \left(\frac{2\pi DJ}{365} \right) \right] \times [(\sin \varphi \sin \delta) + (\cos \varphi \cos \delta \sin \omega_s)] \quad (3)$$

em que I_0 é a constante solar ($= 1367 \text{ Wm}^{-2}$), DJ é o dia juliano começando em 1 de janeiro (para primeiro de janeiro, $DJ = 1$, e para 31 de dezembro, $DJ = 365$), φ é a latitude local (em graus), δ é a declinação solar (em graus) Eq. (4), ω_s é o ângulo solar horário do nascer do sol (em graus) Eq.(5).

$$\delta = 23,45 \times \sin \left[\frac{360(DJ+284)}{365} \right] \quad (4)$$

$$\omega_s = \cos^{-1}[-\tan(\varphi) \times \tan(\delta)] \quad (5)$$

O fotoperíodo (N) é calculado conforme a Eq.(6):

$$N = \frac{2}{15} \times \omega_s \quad (6)$$

A irradiação direta no topo da atmosfera diária (H_{sc}^d , MJm⁻²) é obtida multiplicando a constante solar integrada ($H_{sc} = 4,921$ MJm⁻²) pelo fotoperíodo, (Eq. 7):

$$H_{sc}^d = 4,921 \times N \quad (7)$$

A massa ótica relativa (m_r) foi calculada conforme (Cañada et al., 1993).

4.3.5 Índices estatísticos utilizados para a determinação da eficiência dos métodos

Diversos índices estatísticos podem ser utilizados para avaliar o desempenho dos modelos de estimativa de H_b . Alguns índices são usados aqui para avaliar os modelos de RNA e SVM: Relative Mean Bias Error (rMBE), Relative Root Mean Square Error (rRMSE), coeficiente de determinação (R^2) e d de Willmott (ESCOBEDO et al., 2011). rMBE e rRMSE são expressos em porcentagem da média da medida. As definições destes índices são dadas pelas seguintes equações:

$$rMBE (\%) = 100 \times \frac{\frac{\sum_{i=1}^x (H_E - H_M)}{x}}{\bar{X}} \quad (8)$$

$$rRMSE (\%) = 100 \times \frac{\left[\frac{\sum_{i=1}^x (H_E - H_M)^2}{x} \right]^{\frac{1}{2}}}{\bar{X}} \quad (9)$$

$$R^2 = \left(\frac{\sum_{i=1}^x (H_M - \bar{H}_M)^2 \times \sum_{i=1}^x (H_E - \bar{H}_E)^2}{\sqrt{\sum_{i=1}^x (H_M - \bar{H}_M)^2} \times \sqrt{\sum_{i=1}^x (H_E - \bar{H}_E)^2}} \right)^2 \quad (10)$$

$$d = 1 - \frac{\sum_{i=1}^x (H_E - H_M)^2}{\sum_{i=1}^x (|H'_E| + |H'_M|)^2} \quad (12)$$

em que: H_E representa os valores estimados, H_M os valores medidos, $|H'_E|$ o valor absoluto da diferença $H_E - H_M$, em que $\bar{H}_M$, representa a média de H_M , $|H'_M|$ representa o valor absoluto da diferença $H_M - \bar{H}_M$ e $\bar{H}_E$ a média dos valores estimados. $\bar{X} \left(= \frac{1}{x} \sum_{i=1}^x H_M \right)$ é o valor médio da medida e x o número de observações. O índice rMBE descreve a tendência média dos valores estimados em superestimar (valores positivos) ou subestimar (valores negativos) com relação as medidas. O valor ótimo de rMBE é 0. Os índices rRMSE e R^2 são frequentemente utilizados. O valor ótimo de rRMSE é 0. Bons modelos devem ter valores baixos de rRMSE e rMBE. O valor de R^2 varia de 0 a 1 e quanto maior seu valor melhor o ajuste do modelo. O índice de concordância de Willmontt “d” indica o quão perto as estimativas estão das medidas correspondentes a linha de comparação 1:1. O valor de “d” igual a 1 corresponde a uma combinação perfeita. Para descobrir qual modelo pode ter um desempenho consistentemente alto, (GUEYMARD:RUIZ-ARIAS, 2015), elaboram uma série de critérios para avaliação de modelos de H_b . Diferentes intervalos de rRMSE são definidos para avaliar a acurácia dos modelos (JAMIESON et al., 1991; HEINEMANN et al., 2012):

Excelente se (excellent if) $\text{rRMSE} < 10\%$;

Bom se (good if) $10\% \leq \text{rRMSE} < 20\%$;

Aceitável se (fair if) $20\% \leq \text{rRMSE} < 30\%$;

Pobre se (poor if) $\text{rRMSE} \geq 30\%$.

Neste trabalho, estes intervalos servirão como parâmetros para classificar a precisão dos modelos na estimativa de H_b e definir o modelo mais recomendado. Aqui será considerado erros baixos para $|\text{rMBE}| < 10\%$. A importância de utilizar critérios decorre do fato de que qualquer erro na estimativa de recursos solares podem ter um efeito negativo sobre a simulação de potência, a capacidade de banco e o sucesso global de grandes projetos de energia solar (GUEYMARD:RUIZ-ARIAS, 2015).

4.3.6 Descrição do local de estudo e da base de dados

Os dados utilizados neste trabalho foram medidos na Estação Radiométrica situada na Faculdade de Ciências Agronômicas da UNESP de Botucatu (latitude 22,85°S, longitude 48,45°W e altitude 786m). A região possui elevado gradiente de altitude, entre 400 a 500 m na região mais baixa (depressão periférica) e entre 700 a 900 na região serrana (Planalto Ocidental). Com bioma de cerrado e de mata atlântica, possui clima temperado quente (mesotérmico), verão quente e úmido com elevada precipitação e inverno seco (ESCOBEDO et al., 2011). Com temperatura do ar média anual de $20,46 \pm 2,21^{\circ}\text{C}$, o mês mais quente é fevereiro ($23,216 \pm 1,20^{\circ}\text{C}$) e o mais frio julho ($17,16 \pm 1,33^{\circ}\text{C}$). A umidade relativa do ar varia de $62,61 \pm 8,88\%$ (agosto) a $76,26 \pm 8,24\%$ (fevereiro). A precipitação média anual acumulada é de 1.494,10 mm. O período mais chuvoso ocorre de outubro a março e o menos chuvoso entre abril e setembro. No período chuvoso, a precipitação é causada pela Zona de Convergência do Atlântico Sul (ZCAS). No período seco, a precipitação é do tipo frontal originada do encontro das massas de ar frias e secas vindas da região sul com as massas quentes e úmidas, da região sudeste.

Neste estudo H_b foi medida por um pireliômetro da Eppley NIP acoplado a um rastreador solar ST3 da Eppley, a H_G por piranômetro Eppley PSP e a insolação solar por um heliógrafo convencional do tipo Campbell-Stokes. A temperatura do ar foi obtida por termômetro de máximo e de mínimo, a umidade relativa do ar foi medida pelo higrôgrafo, instalados dentro de um abrigo meteorológico. As medidas são do período de fevereiro de 1996 a dezembro de 2008. Os sensores são verificados diariamente para reposicionamento quando necessário. As medições em irradiância são realizadas a cada 5 s, com armazenamento de média a cada 5 min. As irradiâncias (Wm^{-2}) são integradas em intervalos de tempo específicos para obtenção das irradiações (MJm^{-2}) a partir de programas para análise de radiação (CHAVES:ESCOBEDO, 2000). Os radiômetros são aferidos anualmente, para diferentes coberturas de céu, pelo método comparativo sugerido pela Organização Mundial de Meteorologia (WMO, 2008). Como não há nenhum procedimento definitivo, ideal ou amplamente aceito para melhor controle de qualidade dos dados de irradiação, cada centro de pesquisa tipicamente adota seu próprio método, que implica em alguns podem ser mais rigorosos que outros (GUEYMARD:RUIZ-ARIAS, 2015).

4.4 RESULTADOS E DISCUSSÃO

Inicialmente, a RNA e SVM foram treinadas e, posteriormente utilizadas na estimativa de H_b . Nos modelos de RNA foram considerados os valores: Learning rate = 0,3; momentum = 0,2 e number of iteration = 500. As camadas ocultas (hiddenlayers) foram testadas variando de 1 – 10, porém o valor padrão do WEKA foi adotado pelo melhor ajuste encontrado. No WEKA o padrão das camadas ocultas é definido como “a” = [(variáveis de entrada + classes)/2]. No treinamento da SVM a seleção e a utilização apropriada da função de Kernel têm grande precisão sobre os dados modelados e sobre os modelos de SVM (CHEN et al., 2011). Os três parâmetros ajustados da RBF foram: C , γ e ε . Na seleção destes, fixou-se o valor de ε em 0,005 e foram realizados inúmeros ensaios com diferentes combinações para C e γ . Então, os melhores valores e utilizados foram: $C = 50$, $\gamma = 0,2$ e $\varepsilon = 0,005$. Duas combinações dos parâmetros da RBF sugeridos para a H_G foram testadas: $C = 100$; $\gamma = 0,3$; $\varepsilon = 0,001$ (RAMEDANI et al., 2014) e $C = 400$; $\gamma = 0,01$; $\varepsilon = 0,4$ (MOHAMMADI et al., 2015b), mas os valores de R^2 gerados para os modelos com estes parâmetros foram menores que os obtidos neste trabalho para H_b (Tabela 2).

Tabela 2. Coeficientes de determinação (R^2) dos modelos de RNA e SVM.

Horário				Diário			
RNA	R^2	SVM	R^2	RNA	R^2	SVM	R^2
HRNA1	0.53	HSVM1	0.52	DRNA1	0.49	DSVM1	0.55
HRNA2	0.88	HSVM2	0.88	DRNA2	0.88	DSVM2	0.88
HRNA3	0.88	HSVM3	0.88	DRNA3	0.88	DSVM3	0.88
HRNA4	0.88	HSVM4	0.88	DRNA4	0.88	DSVM4	0.88
HRNA5	0.88	HSVM5	0.88	DRNA5	0.92	DSVM5	0.88
HRNA6	0.88	HSVM6	0.88	DRNA6	0.92	DSVM6	0.88
HRNA7	0.88	HSVM7	0.86	DRNA7	0.69	DSVM7	0.45
**	**	**	**	DRNA8	0.62	DSVM8	0.45

Os valores de R^2 para os 15 modelos (7 horário e 8 diário) para RNA e SVM, correspondem aos valores obtidos seguindo a metodologia Percentage Split do WEKA. Os modelos com os mais baixos valores de R^2 estão destacados em negrito. Os maiores valores de R^2 foram obtidos com os modelos diários. O menor valor de R^2 (0,45) foi para (DSVM7 e DSVM8) e o maior valor de R^2 (0,92) para três modelos (DRNA5, DRNA6 e DSVM5). Nos modelos horários a partir de HRNA2 e

HSVM2 os valores de R^2 são basicamente os mesmos (0,88) para todas as combinações de variáveis. A exceção é para o HSVM7 que tem R^2 (0,86). Na modelagem diária R^2 melhorou com a adição de k_t (DRNA2 e DSVM2), nos modelos seguintes R^2 continua os mesmos valores até a adição da razão de insolação, que melhorou o valor de R^2 para 0,92 nos modelos (DRNA5, DRNA6). Os modelos (DRNA7, DRNA8, DSVM7 e DSVM8) não melhoram os valores de R^2 . Em geral, a similaridade nos valores de R^2 entre (RNA e SVM) e a eficiência na modelagem ganha grande destaque, principalmente quando os parâmetros são ajustados adequadamente e são utilizadas boas medidas (LORENA et al., 2011).

Além de verificar a melhor técnica de aprendizagem de máquina, este estudo também identifica as variáveis climáticas de entrada que têm importância na determinação de H_b . Para isso, foi utilizado o algoritmo J48 do WEKA como classificador das variáveis de entrada mais relevantes. Obteve-se nos modelos horário: k_t e H_{sc} ; nos modelos diários: k_t e razão de insolação (n/N). Observa-se que os melhores valores de R^2 são para os modelos com a adição destas variáveis. O k_t e a m_r como as variáveis de entradas mais relevantes na modelagem de H_b é destacada no estudo de (LOPEZ et al., 2005). Nas seções seguintes serão analisados se os melhores resultados coincidem com os modelos que possuem as variáveis de entrada destacadas.

4.4.1 Análise do desempenho dos modelos horário

Com escassez de trabalhos que utilizam AMs na estimativa de H_b (YADAV:CHANDEL, 2014), os resultados aqui limitam-se a poucas comparações com a literatura, mas enfatiza-se, desde já, a importância de explorar e validar novas metodologias de estimativa de H_b para suprir a demanda mundial de informações solarimétricas para diferentes aplicações, incluindo calibração de satélites, mapeamento solarimétrico e recuperando séries históricas. A discussão seguinte enfatiza o significado da seleção adequada das variáveis de entrada para melhorar a precisão da estimativa de H_b por meio da RNA e SVM.

Na avaliação dos resultados com os modelos (HRNA e HSVM) foram utilizados os índices de validação ($rMBE$, $rRMSE$ e d) (Figuras 1a, b, c). Todos os modelos superestimam ($rMBE > 0$). Os valores de $rMBE$ com RNA são em média 10,75% maiores que os obtidos com SVM. Os resultados de $rRMSE$ e “ d ” com os modelos de SVM são respectivamente, menores e maiores, que os gerados com os

modelos de RNA. Os HSVM1 e HRNA1 apresentaram precisão pobre na estimativa de H_b^h , ($rRMSE = 50,05\%$) com HSVM1 e ($rRMSE = 52,79\%$) com HRNA1. Os valores de $rMBE$ ($20,16\%$ com HRNA1 e $3,51\%$ com HSVM1) e “d” ($0,82$ com HRNA1 e $0,85$ com HSVM1) destes modelos foram os maiores e os menores dos modelos horário, respectivamente. O resultado comprova que apenas H_G^h como variável de entrada não é recomendada para estimar H_b^h , pois, a variação temporal de H_b quando comparada com a H_G é maior (MEYER et al., 2009), portanto, maior a dificuldade para modelarem e baixa será a correlação entre as duas componentes. Então, novas variáveis de entrada são necessárias para gerar modelos e estimativas melhores para H_b (MARQUEZ:COIMBRA, 2011). Por isso, nos modelos seguintes foram adicionadas novas variáveis. Com a adição do kt nos modelos HRNA2 e HSVM2 houve melhora nos valores de $rMBE$, $rRMSE$ e d. Assim, os modelos HRNA2 e HSVM2 estimam H_b^h com $rRMSE = 25,18\%$ e $rRMSE = 19,89\%$, respectivamente, e superestimam as medidas com $rMBE = 14,12\%$ e $rMBE = 1,16\%$ para HRNA2 e HSVM2, respectivamente. O HSVM2 tem melhor “d” ($0,98$) que HRNA2 ($0,96$). O uso de kt como variável de entrada é interessante por ser um indicativo da cobertura de nuvens, elimina o efeito astronômico e evidência o efeito climático.

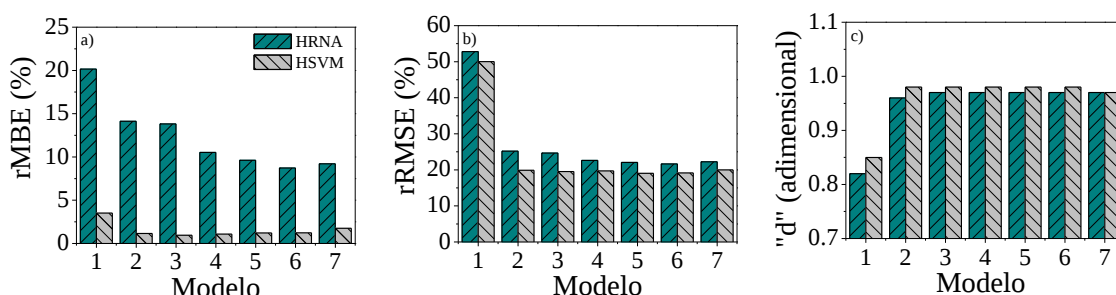


Figura 1. Índices de validação dos modelos RNA e SVM na estimativa da H_b^h . a) $rMBE$, b) $rRMSE$ e c) “d”.

Nos modelos HRNA3 e HSVM3 a H_{sc} pouco melhorou a estimativa. A inserção de H_0 (HRNA4) melhorou $rRMSE$ em $\approx 2\%$ quando comparado com o HRNA3 e em $\approx 2,5\%$ com relação ao HRNA2. Os modelos HRNA5 e HRNA7 têm resultados similares ao HRNA4. Assim, a m_f como variável de entrada nos modelos de redes neurais pode ser utilizada na estimativa de H_b^h . O melhor desempenho entre os modelos das redes é para HRNA6: $rMBE = 8,73\%$, $rRMSE = 21,64\%$ e $d = 0,97$. Com a inserção de novas variáveis houve melhora gradativa nos modelos de redes neurais.

Contudo, não existe diferença significativa nos valores de rRMSE e “d” para os modelos HRNA4 – HRNA7. Com exceção de HRNA1, as combinações das variáveis de entrada dos modelos de rede neural podem ser utilizadas na estimativa de H_b^h .

Nos modelos de SVM (HSVM2 – HSVM7), a adição de novas variáveis de entrada não melhorou significativamente a estimativa de H_b^h . Nestes modelos, os valores de rMBE ficaram entre 0,95 e 1,75%, rRMSE entre 19,06 e 19,98% e $d > 0,97$. Portanto, a SVM mostra melhor estabilidade na modelagem de H_b^h que a RNA. Algumas modificações nos parâmetros da função RBF (C e γ) foram realizadas para verificar se a estimativa para os modelos separadamente melhorava. Como não houve ganho na modelagem, os valores já destacados continuam fixos. Assim, a estabilidade ocorre quando é definido corretamente o valor do parâmetro C e então, os valores discrepantes dos parâmetros são facilmente supridos. Portanto, os resultados confirmam que a SVM é capaz de produzir resultados precisos (CRISTIANINI:TYLOR, 2000), tem maior capacidade de generalização e potencial de rastrear dados históricos para melhorar a previsão futura de séries (HAN et al., 2007).

Os resultados deste estudo assemelham-se aos encontrados por Eissa et al. (2013), que utilizaram RNA para estimar H_b^h por imagem de satélite e obtiveram (rRMSE = 26,10% e rMBE = -6,01%). Linares-Rodriguez et al., (2015) mostram que na literatura estimativas de H_b^h a partir de satélites têm em média rRMSE ~35,00%. A partir de um método proposto, Fernández-Peruchena et al., (2014), geraram uma série climática de H_b com valores a cada 1 min e encontraram erro médio > 30,0%. Gerando série sintética a cada 5 min de H_b , Grantham et al., (2013), obtiveram rMBE = -0,40% e rRMSE = 16,30%. A diferença dos resultados local para os supracitados está relacionada à metodologia adotada. O método utilizado pelos autores apresenta condições de céu em intervalos com incremento de 0,01, que é diferente da metodologia local. Os resultados locais são melhores que os obtidos por Polo et al., (2011), que relataram rRMSE = 31,0% na geração de uma série de H_b a cada 10 min a partir de uma determinada série horária. Linares-Rodriguez et al., (2015) estimaram H_b^h por satélites, utilizando cinco modelos de RNA, e obtiveram rRMSE entre 24,23 e 37,60%. Isto mostra que os modelos locais e as variáveis utilizadas têm melhor desempenho na estimativa de H_b^h . Esta diferença deve-se, principalmente, aos maiores erros esperados quando são utilizados os produtos de satélites. Os valores de rMBE encontrados por Linares-Rodriguez et al., (2015) assemelham-se aos obtidos localmente.

Os gráficos de caixa mostram os erros entre as medidas e as estimativas de H_b^h com os modelos HRNA (Figura 2a) e HSVM (Figura 2b). Dentro de cada caixa, a marca central mostra a média de todos os valores de erro. Por padrão, a caixa é determinada pelos percentis 25 e 75. Os bigodes são determinados pelos percentis 5 e 95. Com exceção dos modelos HRNA1 e HSVM1, os demais têm caixa estreita, ou seja, as estimativas são mais confiáveis. Estes resultados concordam com o desempenho dos modelos analisados com os índices de validação.

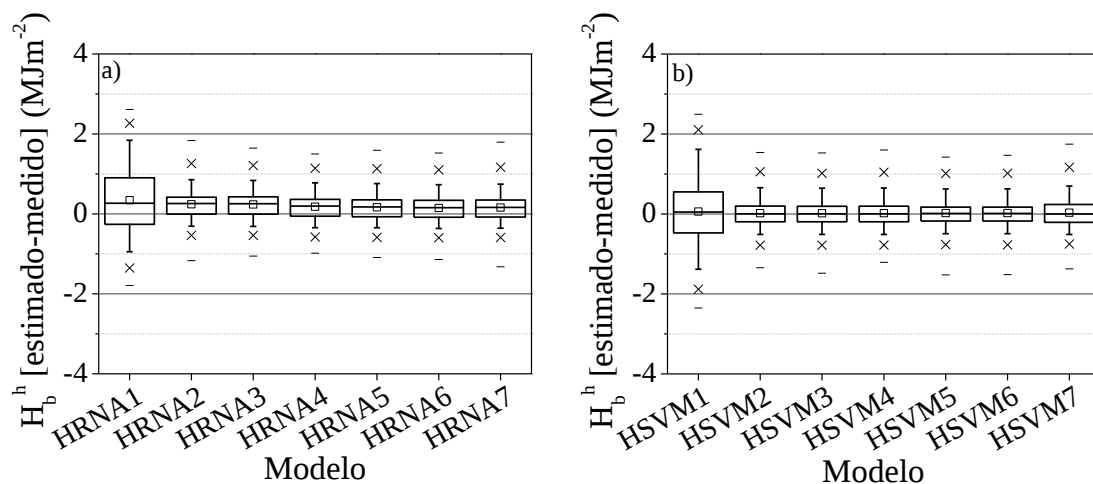


Figura 2. Box plots de erros dos quatorze modelos horário. a) HRNA e b) HSVM.

Com base no rRMSE os modelos HRNA1 e HSVM1 têm precisão pobre, os modelos HRNA2 – HRNA7 têm precisão aceitável e os modelos HSVM2 – HSVM7 precisão boa. Os resultados observados em Botucatu são favoráveis para a aplicação da SVM e RNA. No entanto, mostram que SVM possui melhor desempenho na estimativa de H_b^h que RNA.

A dispersão entre H_b^h medida e estimada, pelas modelos que tiveram a pior precisão e o com melhor precisão são mostrados nas Figuras (3a-d). Os modelos HSVM1 (Figura 3a) e HRNA1 (Figura 3b) tiveram a maior dispersão. A elevada dispersão para estes modelos é causada pelo efeito climático de H_b^h ser maior que o efeito astronômico. Ou seja, a variabilidade temporal de H_b^h , numa atmosfera com nuvens, é diferente da variação temporal de H_G^h sob a mesma atmosfera. As nuvens são as principais responsáveis pela redução de H_G , na ausência das nuvens os aerossóis são os principais responsáveis pela redução de H_b^h (GUEYMARD, 2012), que dependendo da concentração de aerossóis na atmosfera, a redução para H_b pode varia de 30% a 100% e para H_G é inferior a 10%. O vapor d'água na atmosfera, atuando como absorvedor, também reduz significativamente os valores de H_b . Devido à baixa relação

entre o aumento e a redução destas radiações, a H_G sozinha não deve ser aplicada diretamente na estimativa de H_b (LAW et al., 2014). Os modelos HRNA6 (Figura 3c) e HSVM3 (Figura 3d) tiveram os menores espalhamentos e mostram boa concordância. As curvas dos demais modelos não são mostradas, mas possuem comportamentos similares aos dos modelos supracitados. Por fim, as dispersões evidenciam que para as variáveis de entrada ideais a RNA e a SVM AMs são capazes de modelar a H_b^h .

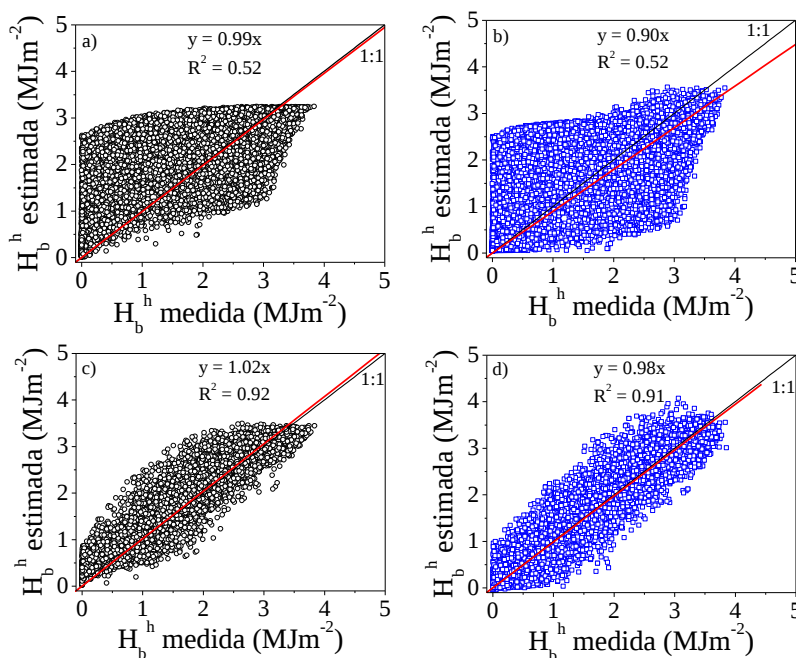


Figura 3. Dispersão dos dados medidos e estimados por RNA e SVM para H_b^h : a) HRNA1, b) HSVM1, c) HRNA6 e d) HSVM3.

4.4.2 Análise do desempenho dos modelos diário

Os índices de validação (rMBE, rRMSE e d) dos modelos diários mostram que a estimativa com SVM é melhor que a estimativa com RNA (Figura 4a, b, c). Os modelos que têm como variável de entrada somente H_G^d , DRNA1 e DSVM1, estimaram com rRMSE = 26,91% e rRMSE = 22,08%, respectivamente. Com a inserção de kt, o DRNA2 reduziu em $\approx 4,65\%$ o valor de rRMSE comparado com DRNA1. O modelo DRNA3 não tem melhora na estimativa com a adição de H_{sc}^d ao ser comparado com DRNA2. O DRNA4 melhorou a estimativa (rMBE = 2,14%, rRMSE = 18,64% e d = 0,97) com a variável H_0 adicionada. Quando se compara DRNA4 com DRNA1 a redução de rRMSE foi $\approx 8,30\%$, com DRNA2 de $\approx 3,60\%$ e com DRNA3 de $\approx 3,50\%$. Quando comparado o rMBE do DRNA4 com estes modelos, a melhoria é

mais significativa. O modelo DRNA5 mostra $rMBE = 6,41\%$, $rRMSE = 17,44\%$ e $d = 0,98$. Enquanto que, o DRNA6 resultou em $rMBE$, $rRMSE$ e d iguais a $6,93\%$, $17,15\%$ e $0,98$, respectivamente. Nestes modelos (DRNA5 e DRNA6) a inserção da razão de insolação (n/N) melhorou os valores de $rRMSE$ e d , no entanto, aumentou o $rMBE$ em $\approx 4,43\%$ em comparação com DRNA4. Portanto, a combinação das variáveis de DRNA4 é a mais indicada.

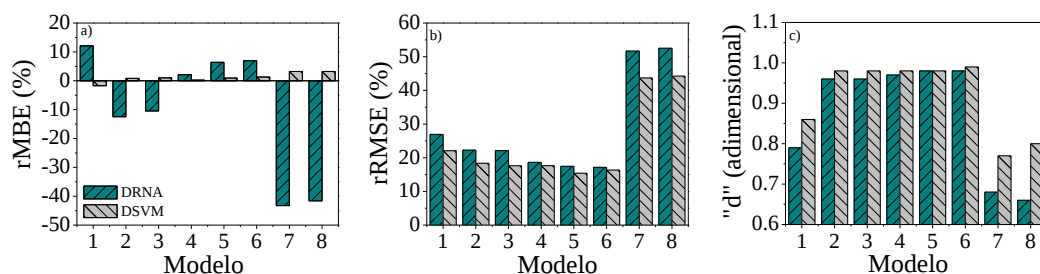


Figura 4. Índices de validação dos modelos RNA e SVM na estimativa de H_b^d . a) $rMBE$, b) $rRMSE$ e c) “ d ”.

As piores estimativas são para os modelos baseados em T_{max} , T_{min} , T_{med} , UR e n . Assim, o DRNA7 e DRNA8 resultaram em $rRMSE$ iguais a $51,67\%$ e $52,51\%$, respectivamente. Para estes modelos os valores de $rMBE$ ($-43,27\%$ com DRNA7 e $-41,60$ com DRNA8) e d ($0,68$ com DRNA7 e $0,66$ com DRNA8) são os maiores e menores, respectivamente. Portanto, as variáveis utilizadas não são recomendadas para estimativa de H_b^d . Os modelos DRNA2, DRNA3, DRNA7 e DRNA8 subestimaram as medidas, enquanto que os demais superestimaram.

Os modelos de SVM tem menores $rMBE$ que os modelos de RNA. O DSVM1 é o único modelo de SVM a subestimar ($rMBE = -1,74\%$). A adição de novas variáveis de entrada melhorou a precisão dos modelos de SVM. Os modelos DSVM2 – DSVM6 têm $rMBE$ no intervalo de $0,24$ a $1,29\%$ ($média = 0,87 \pm 0,39\%$), $rRMSE$ no intervalo de $15,38\%$ a $18,36\%$ ($média = 17,07 \pm 1,19\%$) e $d = 0,98$. Os modelos DSVM7 e DSVM8 são os de SVM com o pior desempenho. Isto é explicado pelas variáveis de entrada. Apesar dos vetores de entrada da SVM serem bastante flexíveis, as variáveis de entrada (T_{max} , T_{min} , T_{med} , UR e n), não foram facilmente incorporadas aos modelos. Sendo H_b fortemente atenuada pelos constituintes atmosféricos, muito provavelmente, a substituição destas variáveis de entrada por informações de aerossóis e vapor d’água resultaria em melhoria na estimativa. Neste

trabalho não são considerados os aerossóis e o vapor d'água por serem variáveis não facilmente disponíveis.

Similarmente aos modelos horários, os DRNA e DSVM tiveram melhores precisões com a inserção de outras variáveis. No caso dos modelos DRNA1 e DSVM1, a explicação para a precisão pobre pode ser resumindo da seguinte maneira: as nuvens na atmosfera resultam em redução dos valores de H_b^d , que causam elevação na irradiação difusa diária (H_D^d). Essa dinâmica na atmosfera faz com que nos dias de céu claro a H_G^d e a H_b^d sejam elevadas. Enquanto que, nos dias de céu nublado a H_b^d tende a zero e a H_G^d é reduzida significativamente a valores menores, mas superiores aos valores de H_b^d devido à componente H_D^d que eleva os valores de H_G^d . Esta dinâmica faz com que a relação entre H_b e H_G seja baixa, que influencia na modelagem. Portanto, os resultados mostram que a inserção de mais variáveis melhora a estimativa de H_b^d com RNA e SVM.

Os gráficos de caixa mostram os erros entre as medidas e as estimativas de H_b^h com os modelos DRNA (Figura 5a) e DSVM (Figura 5b). Dentro de cada caixa, a marca central mostra a média de todos os valores de erro. Por padrão, a caixa é determinada pelos percentis 25 e 75. Os bigodes são determinados pelos percentis 5 e 95. Com exceção dos modelos DRNA1, DRNA7, DRNA8, DSVM1, DSVM7 e DSVM8, os demais têm caixa estreita, ou seja, as estimativas são mais confiáveis. Estes resultados concordam com o desempenho dos modelos analisados com os índices de validação.

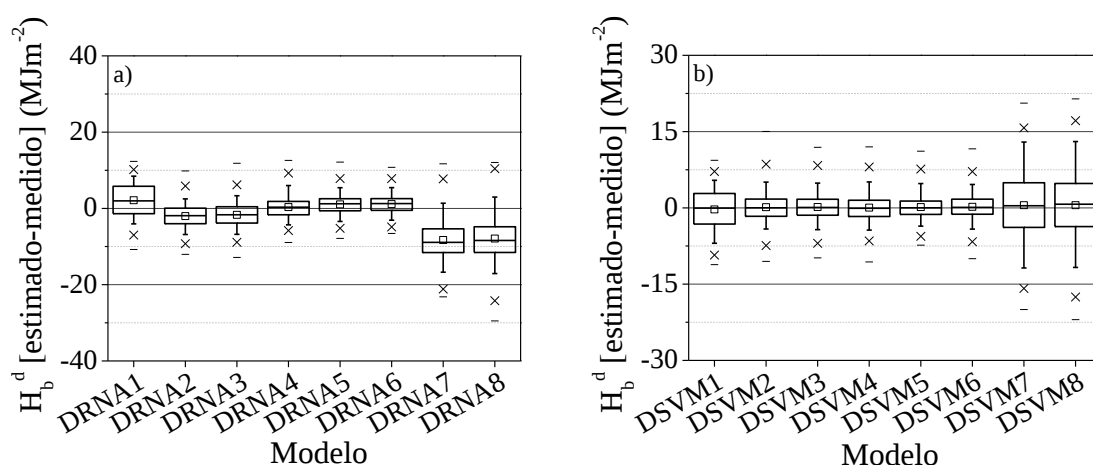


Figura 5. Box plots de erros dos dezesseis modelos diário. a) DRNA e b) DSVM.

Com base em rRMSE os modelos DRNA1, DRNA2, DRNA3 e SVM1 têm precisão aceitável, os modelos DRNA4 – DRNA6 e DSVM2 – DSVM6 têm

precisão boa e os modelos DRNA7 – DNRA8 e DSVM7 – DSVM8 tem precisão pobre. Portanto, a SVM possui melhor desempenho na estimativa de H_b^d que RNA.

Algumas combinações possíveis foram consideradas para encontrar um conjunto de variável de entrada e a técnica de aprendizado mais favorável na estimativa de H_b^d . Oito modelos com diferentes combinações de variáveis de entradas foram estabelecidos. Os valores estimados de H_b^d com RNA e SVM em função das medidas para os melhores e piores modelos são apresentados (Figuras 6a - f).

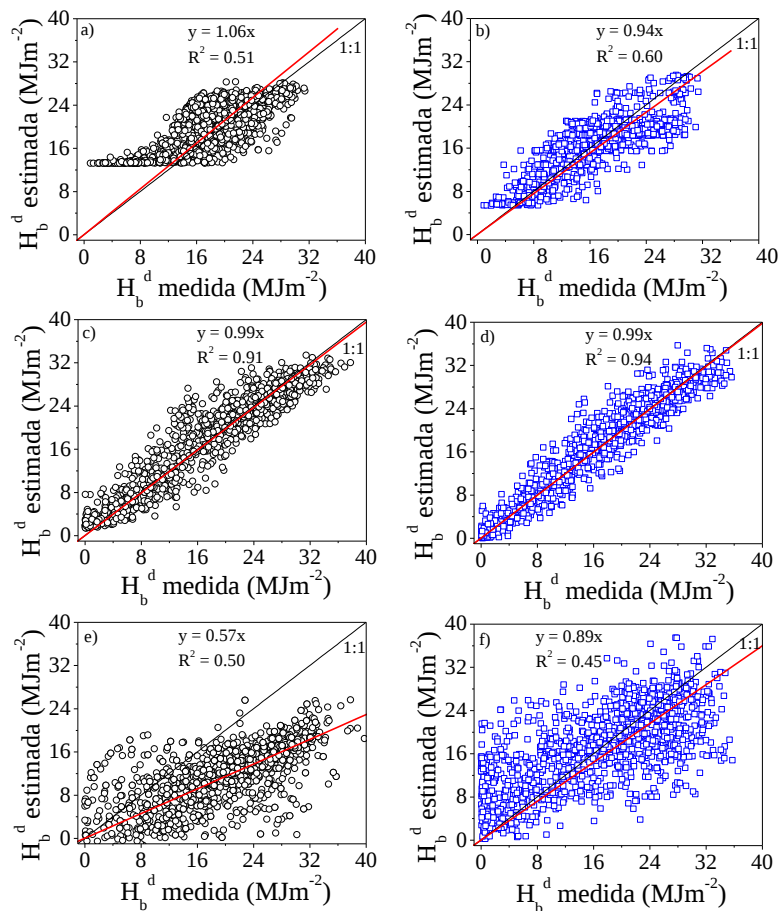


Figura 6. Dispersão dos dados medidos e estimados por RNA e SVM para H_b^d : a) DRNA1, b) DSVM1, c) DRNA4, d) DSVM5, e) DRNA8 e f) DSVM8.

A dispersão com os modelos DRNA1 (Figura 6a) e DSVM1 (Figura 6b) comprovam a baixa correlação entre H_b^d e H_G^d . Os modelos DRNA4 (Figura 6c) e DSVM5 (Figura 6d) mostram que muitos pontos estão ao longo da linha ideal de comparação (1:1) e representa boa concordância entre os valores medidos e os estimados. A capacidade de cada modelo e a técnica para oferecer estimativas precisas está relacionada com a seleção adequada das variáveis de entrada. Por isso, nos modelos

DRNA8 (Figura 6e) e DSVM8 (Figura 6f) muitos pontos estão distantes da linha 1:1. As elevadas superestimativas para DRNA8 e DSVM8 são observadas nas curvas de dispersão.

4.5 CONCLUSÕES

O trabalho apresentado aqui avalia dois métodos de aprendizado de máquinas para estimativa de H_b nas partições, horária e diária, e foram utilizados 13 anos de dados de alta qualidade para treinamento e validação. A RNA e SVM foram introduzidas na no WEKA. Foram utilizados diferentes variáveis de entrada para identificações das que apresentam melhores correlações com H_b . A estimativa de H_b e avaliação de métodos são de grande importância para aplicação na área de conversão de energia solar e na localização de áreas com potencial para instalação de usinas.

Na estimativa de H_b com a SVM, recomenda-se a utilização dos parâmetros: $C = 50$, $\gamma = 0,2$ e $\varepsilon = 0,005$, para a RBF. O trabalho não tinha o interesse de avaliar outras funções de Kernel. Todavia, outros trabalhos podem ser considerados na avaliação da flexibilidade de outras funções na estimativa de H_b . As variáveis de entrada mais relevantes nos modelos foram à H_{sc} , kt e a razão de insolação. Quando existe relação significativa entre as variáveis de entrada e a modelada, a inserção de mais variáveis melhora a estimativa de H_b .

A dispersão pelas duas técnicas de AMs mostra boa concordância para a maioria dos modelos. Portanto, estes resultados evidenciam que ambas as técnicas de AMs são capazes de modelar a H_b nas duas partições de tempo (horária e diária) com resultados satisfatórios.

A partir dos resultados observados em Botucatu, as técnicas de AMs foram capazes de contornar e modelar as diferentes combinações de dados de entrada em todas as condições atmosféricas. Neste trabalho o desempenho das técnicas de AMs não foi comparado com modelos estatísticos ou de transferência radiativa. A comparação dos resultados é somente entre as duas técnicas. Os modelos HRNA1 e HSVM1 tiveram as piores estimativas dos modelos horário. Os modelos HRNA6 e HSVM3 têm as melhores estimativas. Na modelagem diária, os modelos DRNA8 e DSVM8 tiveram as piores estimativas. Enquanto que, as melhores estimativas são observadas para os modelos DRNA4 e DSVM5. Os resultados são favoráveis para a aplicação da RNA e SVM. Em geral, a SVM possui melhor precisão na estimativa que

RNA e é indicada como a primeira opção. A exploração e validação de novas metodologias de estimativa para suprir a demanda de informações solarimétricas para diferentes aplicações são de grande importância em estudos atuais.

Os resultados são importantes para serem utilizados para comparação e calibração de dados de satélites, melhorando a resolução temporal e espacial. Os modelos podem ser aplicados a outros locais onde existem outras variáveis medidas, recuperando da série histórica ou conjunto de dados. Trabalhos futuros devem ser considerados para análise de outras variáveis meteorológicas na modelagem de H_b . Análise das técnicas para intervalos de tempo menores é interessante para projetos solares. A fim de determinar medidas precisas, a validação da SVM em outros locais torna-se interessante. A avaliação da RNA e da SVM na estimativa de outras irradiações são propostas para trabalhos futuras, já que a demanda de informações solarimétricas aumenta rotineiramente.

4.6 REFERÊNCIAS BIBLIOGRÁFICAS

- ALAM, S.; KAUSHIK, S. C.; GARG, S. N. Computation of beam solar radiation at normal incidence using artificial neural network. **Renewable Energy**, v.31, p.1483–1491, 2006.
- AL-RASHEEDI, M.; GUEYMARD, C. A.; ISMAIL, A.; AL-HAJRAF, S. Solar Resource Assessment over Kuwait: Validation of Satellite-derived Data and Reanalysis Modeling. EuroSun 2014 / **ISES Conference Proceedings** (2014).
- Badescu, V.; Gueymard, C. A.; Cheval, S.; Oprea, C.; Baci, M.; Dumitrscu, A.; Iacobescu, F.; Milos, I.; Rada, C. Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. **Renewable Energy**, v.55, p.85-103, 2013.
- BERTRAND, C.; VANDERVEKEN, G.; JOURNEE, M. Evaluation of decomposition models of various complexity to estimate the direct solar irradiance over Belgium. **Renewable Energy**, v.74, p.618-626, 2015.
- BISHOP, M. C. **Neural Networks for pattern recognition**. Oxford University Press, 1995.
- CAÑADA, J.; PINAZO, J. M.; BOSCA, J. V. Determination of Angström's turbidity coefficient at Valencia. **Renewable Energy**, v.3, p.621–626, 1993.

- Chang, C.-C.; Lin, C.-J. **LIBSVM: a library for support vector machines**. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27. 2011. Software available at: <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>.
- CHAVES, M.; ESCOBEDO, J. F. A software to process daily solar radiation data. **Renewable Energy**, v. 19, p. 339-344, 2000.
- CHEN, J.-L.; LIU, H.-B.; WU, W.; XIE, D.-L. Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study. **Renewable Energy**, v.36, p. 413-420, 2011
- CHEN, S.-T.; YU, P.-S. Pruning of support vector networks on flood forecasting. **Journal of Hydrology**, v.347, p.67–78, 2007.
- Cristianini, N.; Tylor, J. S. **An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods**, Cambridge University Press, UK, 2000.
- EISSA, Y.; MARPU, P. R.; GHERBOUDJ, I.; GHEDIRA, H.; OUARDA, T. B. M. J.; CHIESA, M. Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using SEVIRI images. **Solar Energy**, v.89, p.1-16, 2013
- ESCOBEDO, J. F.; GOMES, E. N.; OLIVEIRA, A. P.; SOARES, J. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. **Renewable Energy**, v.36, p.169-178, 2011.
- FERNÁNDEZ-PERUCHENA, C. M.; BLANCO, M.; BERNARDOS, A. Generation of series of high frequency DNI years consistent with RNA annual and monthly long-term averages using measured DNI data. **Energy Procedia**, v. 49, p. 2321 – 2329, 2014.
- FIORIN, D. V.; MARTINS, F. R.; SCHUCH, N. J.; PEREIRA, E. B. **Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares**. Revista Brasileira de Ensino de Física, v.33, p.1309-1320, 2011. [in Portuguese].
- GRANTHAM, A. P.; PUDNEY, P. J.; BOLAND, J. W.; BELUSKO, M. Synthetically Interpolated Five-minute Direct Normal Irradiance. 20th **International Congress on Modelling and Simulation**, Adelaide, Australia, 1–6 December 2013.
- GUEYMARD, C. A. Direct solar transmittance and irradiance predictions with broadband models. Part I: Detailed theoretical performance assessment. **Solar Energy**, v. 74, p. 355–379, 2003.
- GUEYMARD, C. A. Temporal variability in direct and global irradiance at various time scales as affected by aerosols. **Solar Energy**, v.86, p.3544–3553, 2012.

- GUEYMARD, C. A.; RUIZ-ARIAS, J. A. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. **Sol. Energy** (2015), <http://dx.doi.org/10.1016/j.solener.2015.10.010>.
- GUEYMARD, C. A.; RUIZ-ARIAS, J. A. Performance of Separation Models to Predict Direct Irradiance at High Frequency: Validation over Arid Areas. EuroSun 2014 / **ISES Conference Proceedings** (2014).
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMRNA, P.; WITTEN. I. H. **The WEKA Data Mining Software: An Update**; SIGKDD Explorations, v.11, Issue 1, 2009.
- HAN, D.; CHAN, L.; ZHU, N. Flood forecasting using support vector machines. **Journal of Hydroinformatics**, v.9, p. 267-276, 2007.
- HAYKIN, S. **Neural networks: A comprehensive foundation**. 2nd ed. Hamilton: Prentice Hall, 1998, 897p.
- HAYKIN, S. **Redes Neurais: Princípios e prática**. Trad. Paulo Martins Engel. – 2. Ed. – Porto Alegre: Bookman, 2001, 900p. [in portuguese]
- HEINEMANN, A. B.; VAN OORT, P. A. J.; FERNANDES, D. S.; MAIA, A. H. N. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. **Bragantia, Campinas**, v. 71, n. 4, p.572-582, 2012.
- Hsu, C. C.; Chang, C. C.; Lin, C. J **A practical guide to support vector classification**. 2010. <<http://www.csie.ntu.edu.tw/~cjlin/papers/cuide/guide.pdf>>.
- JAMIESON, P. D.; PORTER, J. R.; WILSON, D. R. A test of the computer simulation model ARC-WHEAT1 on wheat crops grown in New Zealand. **Field Crops Research**, v.27, p.337-350, 1991.
- JIAN, Y. Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. **Energy**, v. 34, p.1276-1283, 2009.
- KAUSHINA, N. D. TOMAR, R. K.; KAUSHIK, S. C. Artificial Neural Network model based on interrelationship of direct, diffuse and global solar radiations. **Solar Energy**, v. 103, p.327-342, 2014.
- KHALIL, A. F.; MCKEE, M.; KEMBLOWSKI, M.; ASEFA, T. Basin scale water management and forecasting using artificial neural networks. **Journal of The American Water Resources Association**, v.41, p.195–208, 2005.

- LAM, J. C.; WAN, K. K. W.; LIU, Y. Solar radiation modeling using ANNs for different climates in China. **Energy Conversion and Management**, v.49, p.1080–1090, 2008.
- LAW, E. W.; PRASAD, A. A.; KAY, M.; TAYLOR, R. A. Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting – A review. **Solar Energy**, v.108, p.287–307, 2014.
- LINAREZ-RODRIGUEZ, A.; QUESADA-RUIZ, S.; POZO-VAZQUEZ, D.; TOVAR-PESACOR, J. Na evolutionary artificial neural network ensemble model for estimating hourly direct normal irradiances from Meteosat imagery. **Energy**, v.91, p.264–273, 2015.
- LOPEZ, G.; BATLLES, F-J. Y.; TOVAR-PESCADOR, J. Selection of input parameters to model direct solar irradiance by using artificial neural networks. **Energy**, v.30, p.1675–1684, 2005.
- LORENA, A. C.; JACINTHO, L. F. O.; SIQUEIRA, M. F.; GIOVANNI, R.; LOHMANN, L. G.; CARVALHO, A. C.P.L.F.; YAMAMOTO, M. Comparing machine learning classifiers in potential distribution modelling. **Expert Systems with Applications**, v.38, p.5268–5275, 2011.
- LUNA, A. S.; PAREDES, M. L. L.; OLIVEIRA, G. C. G.; CORRÊA, S. M. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. **Atmospheric Environment**, v. 98, p.98–104, 2014.
- MARQUEZ, R. COIMBRA, C. F. M. Forecasting of global and direct solar irradiance using stochastic learning methods, group experiments and the NWS data-base. **Solar Energy**, v.85, p.746–756, 2011.
- MEYER, R.; BEYER, H. G.; FANSLAU, J.; GEUDER, N.; HAMMER, A.; HIRSCH, T.; HOYER-KLICK, C.; SCHMIDT, N.; SCHWANDT, M. **Towards Standardization Of CSP Yield Assessments**. SolarPACES, Germany, 1–8, 2009.
- MOHAMMADI, K.; SHAMSHIRBAND, S.; ANISI, M. H.; ALAM, K. A.; PETKOVIC, D. Support vector regression based prediction of global solar radiation on a horizontal surface. **Energy Conversion and Management**, v.91, p.433–441, 2015b.
- MOHAMMADI, K.; SHAMSHIRBAND, S.; TONG, C. W.; ARIF, M.; PETKOVIC, D.; CHE, S. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. **Energy Conversion and Management**, v.92, p.162–171, 2015a.

- PANIAGUA-TINEO, A.; SALCEDO-SANZ, S.; CASANOVA-MATEO, C.; ORTIZ-GARCÍA, E. G.; CONY, M. A.; HERNÁNDEZ-MARTÍN, E. Prediction of daily maximum temperature using a support vector regression algorithm. **Renewable Energy**, v.36, p.3054-3060, 2011.
- PIRI, J.; SHAMSHIRBAND, S.; DALIBOR PETKOVIC, D.; TONG, C. W.; REHMAN, M. H. Prediction of the solar radiation on the Earth using support vector regression technique. **Infrared Physics & Technology**, v.68, p.179–185, 2015.
- POLO, J.; L. ZARZALEJO.; MARCHANTE, R.; NAVARRO, A. A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. **Solar Energy**, v.85, p.1164–1170, 2011.
- RAGHAVENDRA, S.; DEKA, P. C. Support vector machine applications in the field of hydrology: A review. **Applied Soft Computing**, v.19, p.372–386, 2014.
- RAMEDANI, Z.; OMID, M.; KEYHANI, A.; SHAMSHIRBAND, S.; KHOSHNEVISAN, B. Potential of radial basis function based support vector regression for global solar radiation prediction. **Renewable and Sustainable Energy Reviews**, v.39, p.1005–1011, 2014.
- SOUZA, JK. L.; NICÁCIO, R. M.; MOURA, M. A. L. Global solar radiation measurements in Maceió, Brazil. **Renewable Energy**, v. 30, p. 1203-1220, 2005.
- TABARI, H.; OZGUR KISI, O.; EZANI, A.; TALAEI, P. H. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. **Journal of Hydrology**, v.444–445, p.78–89, 2012.
- TOMAR, R. K.; KAUSHIKA, N. D.; KAUSHIK, S. C. Artificial neural network based computational model for the prediction of direct solar radiation in Indian zone. **Journal of Renewable and Sustainable Energy** 4, 063146 (2012); doi: 10.1063/1.4772677.
- TONG, K-Y.; ZHU, H. Support vector machine for classification of walking conditions of persons after stroke with dropped foot Hong-yin Lau. **Human Movement Science**, v.28, p. 504–514, 2009.
- TÜFEKCI, P. Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. **Electrical Power and Energy Systems**, v.60, p.126–140, 2014.
- VAPNIK, V. N. **Statistical learning theory**. New York: Wiley, 1998.
- VAPNIK, V. N. **The nature of statistical learning theory**. New York: Springer, 1995.
- Verlag. Mitchell, T. (1997). **Machine learning**. McGraw Hill.

WILLMOTT, C. J. On the validation of models. **Physical Geography**, Delaware, v. 2, p. 184-194, 1982.

WMO – WORLD METEOROLOGICAL ORGANIZATION. **Guide to meteorological Instruments and Methods of Observation**. WMO-nº8, Seventh Edition, 2008, p.1-681, Geneva, Switzerland.

YADAV, A. K.; CHANDEL, S. S. Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model. **Renewable Energy**, v.75, p.675-693, 2015.

YADAV, A. K.; CHANDEL, S. S. Solar radiation prediction using Artificial Neural Network techniques: A review. **Renewable and Sustainable Energy Reviews**, v.33, p. 772–781, 2014.

5. ESTIMATIVA DA FRAÇÃO TRANSMITIDA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL UTILIZANDO SVM COM ALGORITMO SEQUENTIAL MINIMAL OPTIMIZATION.

5.1 RESUMO

Neste trabalho, o algoritmo Sequential Minimal Optimization (SMO_{reg}) foi utilizado na estimativa da fração transmitida da irradiação direta na incidência normal (kt_b). A função Radial Basis Function (RBF) de Kernel foi empregada para regressão. Três modelos estatísticos (#M1, #M2 e #M3) foram desenvolvidos e comparados com três modelos treinados com o algoritmo SMO (SMO1, SMO2 e SMO3). As variáveis de entrada utilizadas foram: (1) transmissividade atmosférica da irradiação solar global (kt) e (2) razão de insolação (n/N) (onde n é a duração do brilho solar e N o fotoperíodo). Para garantir uma análise completa, os modelos foram avaliados na partição horária e diária. Foram utilizadas medidas de 13 anos para região de Botucatu – SP. Na avaliação dos modelos foram usados: *Relative Mean Bias Error* (rMBE), *Relative Root Mean Square Error* (rRMSE), erro percentual relativo (e) e d de Willmott. Os resultados

mostram que o SMO é preciso e tem desempenho melhor que os modelos #M. O modelo horário (SMO1), resultou em $rMBE = 1,79\%$, $rRMSE = 26,28\%$ e $d = 0,97$. No modelo diário (SMO2) o $rMBE$, $rRMSE$ e d foram $1,32\%$, $17,69\%$ e $0,98$, respectivamente. O desempenho dos modelos de SMO é aceitável ($-10\% < e < 10\%$). Por último, o desempenho da SMO na estação chuvosa e seca é analisado para verificar a influência das nuvens, aerossóis e vapor d'água na dispersão das estimativas e o aumento dos erros. No modelo horário a estação chuvosa apresenta $rRMSE \approx 9,00\%$ maior que na estação seca. Na modelagem diária o $rRMSE$ é $\approx 6,10\%$ maior na estação chuvosa que na estação seca. Eliminado o efeito das nuvens, os aerossóis de queimadas agrícolas e ambientais, assim como, o vapor d'água aumenta o espalhamento na estimativa de kt_b em Botucatu. Os resultados mostram que o SMO estima kt_b com desempenho confiável em comparação com os modelos estatísticos e pode ser utilizado na estimativa de kt_b .

Palavras-chave: SVM; transmissividade atmosférica; razão de insolação; WEKA.

5.2 INTRODUÇÃO

A irradiação global (H_G), irradiação difusa (H_d) e a irradiação direta na incidência normal (H_b) possuem grande importância em estudos climáticos, agrônômicos, na engenharia e como fonte de energia renovável. A H_G é monitorada por muitos centros de pesquisas, no entanto, poucos são os que monitoram rotineiramente H_{Dd} e mais escassas são as medidas de H_b (KOTTI et al., 2014). O conhecimento da variabilidade espacial e temporal de H_b é indispensável na concepção e dimensionamento de projetos para geração de energia solar limpa e renovável, incluindo fotovoltaica, térmica e as tecnologias concentradas (BLANC et al., 2015). O aumento na procura por medidas confiáveis de radiação solar é resultante do desenvolvimento de novas tecnologias para o uso de recursos energéticos renováveis, principalmente, em substituição dos combustíveis fósseis. Isto é impulsionado pela demanda mundial de energia, pelos riscos ambientais de muitos países e pela mudança no clima da terra. Nos últimos anos têm-se buscado meios de se obter H_b nos locais onde não existem estações solarimétricas para geração de curvas anuais ou de mapas.

Mapas solares podem fornecer informações do potencial solar e da variabilidade anual de H_b nas diferentes regiões. Como a H_b é alterada pela profundidade ótica do aerossol, vapor d'água e fortemente afetada pela dinâmica de nuvens (GRANTHAM et al., 2013), dados de alta qualidade são de grande relevância para diminuir erros na simulação de projetos de usinas solares (Fernandez-Peruchena et al., 2015), permitindo uma modelagem precisa e confiável.

Nos locais onde não existem estações solarimétricas com sensores monitorando H_b , a modelagem torna-se uma alternativa na estimativa e avaliação dos recursos solares. Para isso, várias metodologias têm sido utilizadas para estimar diretamente ou indiretamente H_b . Liu e Jordan (1960) propuseram a relação entre a fração transmitida da irradiação direta na incidência normal (kt_b) e a transmissividade atmosférica da irradiação global (kt) para obter indiretamente H_b .

A relação proposta por Ångström (1924) tem sido utilizada para correlacionar kt_b com a razão de insolação (n/N). Estas correlações são modificadas e utilizadas na obtenção de H_b nas diversas partições de tempo. Maxwell (1987) propôs um modelo que correlaciona kt_b e kt . Perez et al., (1990) obtiveram melhores estimativas de H_b ao modificaram o modelo de Maxwell.

Louche et al., (1991) sugeriram um modelo polinomial que estima H_b a partir de kt . Fernandez-Peruchena et al., (2014), desenvolveram um método para estimativa horária (H_b^h) a partir de medidas diárias da irradiação solar global. Padovan et al., (2014) estimaram H_b com diferentes modelos estatísticos para instalações de sistemas de concentradores solares e obtiveram resultados com boa precisão. A fim de obter resultados precisos e os melhores locais para instalação de usinas solares, (BEHAR et al., 2015) analisaram 17 modelos de transferência radiativa na Argélia.

Modelos de transferência radiativa, a partir de satélites, têm sido desenvolvidos e utilizados, no entanto, a resolução não é suficiente para avaliar sistemas solares (BEYER et al., 2010; VIANA et al., 2011), pois não produzem boa precisão na estimativa e os dados de satélites não estão facilmente disponíveis.

Nos últimos anos, Máquina de Vetores de Suporte (SVM) ganharam grande destaque na modelagem solar. A SVM tem sido utilizada numa série de estudos para estimar a H_G e tem mostrado ser mais eficiente que os modelos empíricos (CHEN et al., 2013; CHEN:LI, 2014; PIRI et al., 2015). Técnicas de aprendizado de máquinas (AM) foram implementadas em ferramentas computacionais

para análise de dados. Dentre elas, destaca-se a *Waikato Environment for Knowledge Analysis* (WEKA), que tem seu conjunto de algoritmos de AM amplamente utilizados com sucesso em diversas áreas (LORENA et al., 2011; YADAV:CHANDEL, 2015). Entre os algoritmos disponíveis para a SVM, o Sequential Minimal Optimization (SMO) é destaque por resolver problemas complexos com programação quadrática sequencial, possuindo aplicação nas diferentes áreas de pesquisas (PLATT, 1998). Zhang et al., (2009) verificaram que SMO é eficiente para resolução de problemas de seleção probabilística de carteira (portfólio). Oliveira (2006) mostra superioridade do SMO quando comparado com outras técnicas de AM. Zhang et al., (2011) mostraram que o custo de transação em forma de V (*V-Shaped transaction cost*) pode ser contabilizado com SMO. Apesar do sucesso da aplicação em determinadas áreas, o SMO é um algoritmo de regressão que não tem sido avaliado na modelagem solar e, principalmente, nos problemas de predição de H_b .

Neste sentido, busca-se neste trabalho que o SMO pode ser aplicado com sucesso na estimativa de kt_b a partir de medidas de H_G e n , através do WEKA. As relações neste trabalho são de $kt_b \times kt$ (horário e diário) e $kt_b \times n/N$ (diário). O desempenho do SMO é comparado com o desempenho dos modelos estatísticos com as mesmas variáveis de entrada, ajustados com dados locais. Este trabalho também discute a validação dos modelos com duas bases de dados, aqui definidas de ano típico (AT) e ano atípico (AAT). O trabalho está organizado da seguinte forma: na metodologia são mostrados os dados utilizados, o local, descrição da SVM, descrição de SMOreg, dos modelos estatísticos, descrição da obtenção da base de validação e os critérios de validação. O desempenho dos modelos é apresentado nos resultados e discussão. Um estudo de caso é realizado para verificar o desempenho do SMO em duas estações distintas: chuvosa e seca, destacando o efeito das nuvens, aerossóis e vapor d'água.

5.3 MATERIAL E MÉTODOS

5.3.1 Local de estudo e dados

Neste trabalho, os dados utilizados foram da cidade de Botucatu, localizada no Estado de São Paulo – Brasil. Botucatu é a sexta mais populosa cidade da região Centro-Oeste Paulista (138.000 habitantes) e possui área total de 1.483,642 km².

A Estação de Radiometria solar está situada na Faculdade de Ciências Agrônômicas da UNESP/Botucatu (latitude 22,85°S; longitude 48,45°W e altitude 786m). A Figura 1 ilustra a localização de Botucatu e da estação radiométrica. Botucatu é a única cidade da região Centro-Oeste Paulista que possui medidas confiáveis de H_b de longo prazo. Devido a problemas técnicos no pireliômetro, a estação parou de monitorar H_b em 2009. Mas continua monitorando rotineiramente outras radiações (H_G e H_d , por exemplo). Neste sentido, a estimativa de H_b usando novas técnicas é de grande importância para ampliação da série histórica e recuperação de dados.

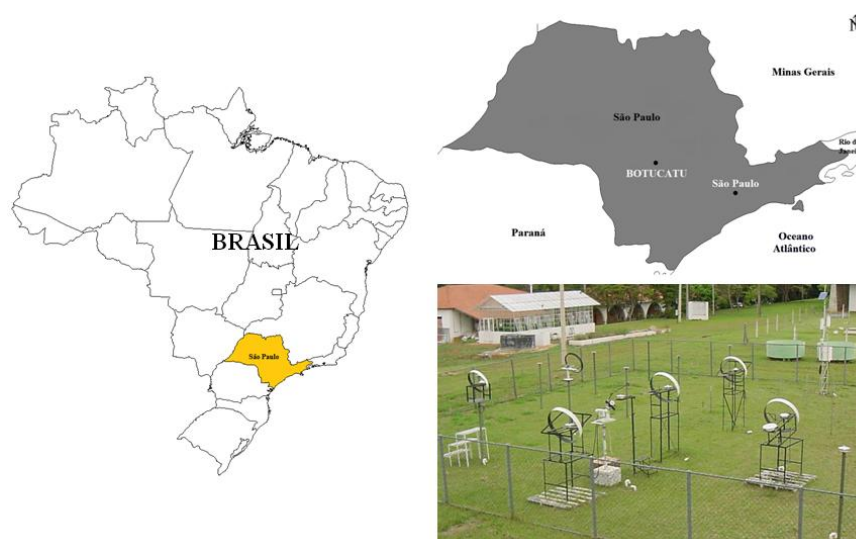


Figura 1. Localização da estação radiométrica.

Com bioma de Cerrado e Mata Atlântica, Botucatu possui clima temperado quente (mesotérmico), verão quente e úmido com elevada precipitação e inverno seco (Escobedo et al., 2011). Com base numa série climática de 38 anos (1971 a 2009), a precipitação média anual acumulada é de 1.494,10 mm, com chuvas mensais entre $40,12 \pm 41,76$ mm (junho) e $270,49 \pm 110,36$ mm (janeiro). A umidade relativa do ar varia de $62,61 \pm 8,88\%$ (agosto) a $76,26 \pm 8,24\%$ (fevereiro), com média anual de $69,93 \pm 4,44\%$. A temperatura do ar varia de $23,21 \pm 1,20^\circ\text{C}$ em fevereiro a $17,16 \pm 1,33^\circ\text{C}$ em julho e temperatura do ar média anual de $20,46 \pm 2,21^\circ\text{C}$. O período mais chuvoso ocorre de outubro a março (primavera-verão) e o menos chuvoso entre abril e setembro. No período chuvoso, a precipitação é causada pela Zona de Convergência do Atlântico Sul (ZCAS). No período seco, a precipitação é do tipo frontal originada do encontro das massas de ar frias e secas vindas da região sul com as massas quentes e úmidas, da região sudeste do Brasil.

Para uma modelagem mais precisa, o uso de uma longa série climática é de grande importância na área solar. Neste estudo, foram utilizados dados de irradiação direta na incidência normal (H_b), irradiação global (H_G) e duração do brilho solar ou insolação solar (n), período de fevereiro de 1996 a dezembro de 2008. A H_b foi obtida por pireliômetro da Eppley NIP acoplado a um rastreador solar ST3 da Eppley, por sua vez, H_G foi monitorada por piranômetro Eppley PSP e n foi obtido por heliógrafo convencionais do tipo Campbell-Stokes. Os dados de H_b e H_G foram separados nas partições de tempo horária e diária. A n foi utilizada na modelagem diária. Os sensores na estação radiométrica são aferidos anualmente pelo método comparativo sugerido pela Organização Mundial de Meteorologia (WMO, 1981).

5.3.2 Máquina de Vetor de Suporte (SVM)

A SVM é um método de Aprendizagem de Máquina supervisionado derivado da teoria de aprendizagem estatística, que analisa bancos de dados e reconhece padrões. Este método é baseado no princípio da minimização do risco estrutural, que tenta minimizar um limite superior de erro de generalização em vez de minimizar o erro de treinamento local (CHEN et al., 2015). A SVM é amplamente utilizada na classificação, análise de regressão e previsão de problemas complexos. O princípio fundamental é fazer a correlação por meio do mapeamento não-linear dos conjuntos de dados de entradas para um espaço de alta dimensão, usando um conjunto de funções de Kernel (VAPNIK, 1995; VAPNIK, 1998).

Neste trabalho o algoritmo *Sequential Minimal Optimization* (SMO) é utilizado para regressão. SMO é um algoritmo simples que rapidamente soluciona e resolve problema de SVM com programação quadrática sequencial de menor otimização possível com dois multiplicadores de Lagrange (SMOLA: SCHÖLKOPF, 2004). Os multiplicadores de Lagrange devem obedecer a uma restrição de igualdade linear. Em cada etapa, SMO escolhe dois multiplicadores de Lagrange para otimização em conjunto e atualiza o SMO para os novos valores ótimos (ZHANG et al., 2009).

Pela eficiência computacional, simplicidade e adaptação para otimização de problemas mais complexos, neste estudo a função a Função de Base Radial (RBF) de Kernel é a escolhida para regressão com SMO. Um Kernel K é uma função que recebe dois pontos x_i e x_j do espaço de entradas e computa o produto escalar desses

dados no espaço de características e a utilidade dos Kernels está, portanto, na simplicidade de seu cálculo e em sua capacidade de representar espaços abstratos (LORENA:CARVALHO, 2007). Na utilização da RBF, primeiramente é necessário ajustar adequadamente os parâmetros C (custo), γ (gamma) e ε (épsilon). Os parâmetros C e γ são mutuamente dependentes, baixos valores de C produzem um AM com má aproximação e valores muito altos geram um aprendizado de máquina mais complexo (RAGHAVENDRA:DEKA, 2014). Os valores de C e γ , são testados empiricamente e aquele com a melhor precisão na validação-cruzada são selecionados. Depois da escolha dos melhores valores de C e γ o conjunto de dados é treinado para geração do modelo final. Já o valor de ε controla como a função vai ajustar os dados de treinamento. Foi utilizado a *Waikato Environment for Knowledge Analysis* (WEKA) para treinamento do algoritmo SMO.

A WEKA é um conjunto de algoritmos de AM para tarefa de mineração de dados, disponível: <<http://www.cs.waikato.ac.nz/~ml/weka/>> (Hall et al., 2009). Fornecendo ao usuário uma programação em Java, o WEKA contém ferramentas para pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização. Foram estabelecidas correlações entre kt_b com kt e kt_b com (n/N) (Tabela 1). Maiores detalhes sobre mineração de dados e o uso WEKA são encontrados em (Witten et al., 2011).

Tabela 1. Modelos de SMO.

Partição	Modelo	Relação
Horário	SMO1	kt_b^h vs kt^h
Diário	SMO2	kt_b^d vs kt^d
Diário	SMO3	kt_b^d vs $(n/N)^d$

5.3.3 Modelos Estatísticos

Os modelos estatísticos foram desenvolvidos conforme metodologias existentes na literatura e são baseados em Liu e Jordan (1960) e Ångström (1924). O primeiro modelo utiliza relações entre kt_b com a kt . Estas frações estão representadas pelas relações: $kt_b^x = H_b^x/H_{sc}^x$ e $kt_t^x = H_G^x/H_0^x$, onde X representa as partições horária (h) e diária (d), H_b^x é a irradiação direta na incidência normal e H_{sc}^x é a constante solar integrada no topo da atmosfera, H_G^x é a irradiação global e H_0^x é a

irradiação solar no topo da atmosfera. As equações de estimativa foram ajustadas por regressão polinomial entre as correlações kt_b^x e kt^x . A fração transmitida diária kt_b^d foi relacionada com a razão de insolação ($r = n/N$).

5.3.4 Geração dos modelos estatísticos

As Figuras 2 (a, b) apresentam a correlação entre kt_b e kt utilizadas para geração dos modelos estatísticos a partir de uma regressão. Foi utilizada uma série climática de 13 anos (1996 a 2008), na partição de dados horária utilizaram-se 47.217 horas (Figura 2a) e na diária 4.304 dias (Figura 2b). Os dados foram analisados através de uma curva de dispersão e os valores inconsistentes foram excluídos com metodologia similar a (JOURNÉE:BERTRAND, 2011). Na curva existe uma gama de valores de kt_b relacionado a único valor de kt . Este efeito é devido às diferentes coberturas de nuvens. Nas regressões para as correlações horárias e diárias foram feitos um ajuste com polinômio do 4°. O uso de polinômio de grau superior a 4 não resultou em quaisquer melhoria no coeficiente de correlação no ajuste dos modelos. As melhorias ocorreram a partir da segunda casa decimal, que torna-se insignificativa.

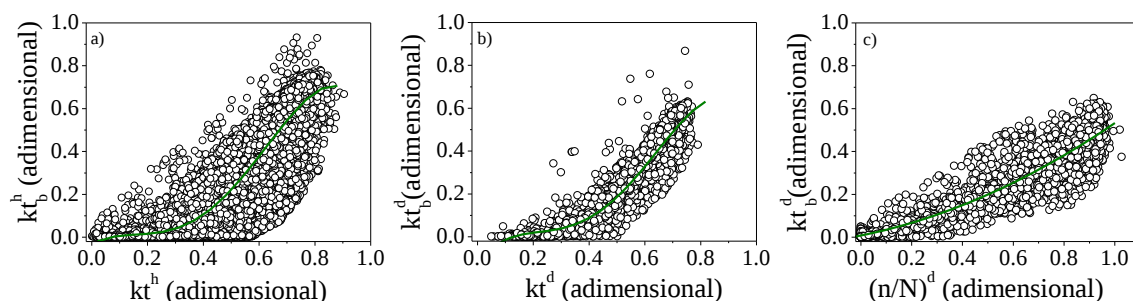


Figura 2. a) $kt_b^h \times kt^h$; b) $kt_b^d \times kt^d$; c) $kt_b^d \times n/N$.

A Figura 2 (c) apresenta a correlação entre kt_b^d e $(n/N)^d$. Neste modelo foi utilizado 4.042 dias. As maiores dispersões são para valores de $(n/N)^d > 0,50$ e um único valor de $(n/N)^d$ pode corresponder a diferentes valores de kt_b^d . A queima do heliograma pode ser a principal causa da dispersão. O modelo foi gerado através de ajuste polinomial do 2° grau. Funções de grau superior não resultaram em melhorias no ajuste do modelo. Os modelos estatísticos gerados e seus respectivos coeficientes de determinação (R^2) são mostrados na Tabela 2.

Tabela 2. Modelos estatísticos gerados.

	Modelo	R ²
#M1	$kt_b^h = -0,035 + 0,775 \times kt - 4,805 \times kt^2 + 12,503 \times kt^3 - 7,906 \times kt^4$	0,75
#M2	$kt_b^d = -0,123 + 1,768 \times kt - 8,213 \times kt^2 + 16,617 \times kt^3 - 9,584 \times kt^4$	0,81
#M3	$kt_b^d = 0,007 + 0,239 \times (n/N) + 0,285 \times (n/N)^2$	0,74

5.3.5 Base de dados para o treinamento e validação dos modelos

Para validação, aqui foi utilizada duas base de dados das irradiações (horária e diária) denominadas por ano típico (AT) e ano atípico (AAT), as quais foram selecionadas da série. A seleção do AT e AAT foram através de análises estatísticas onde foi comparado para cada mês do ano o valor da irradiação média horária entre – anos considerando o intervalo de variação do desvio padrão da média total com o valor da irradiação média de cada ano. A Figura 3 apresenta o exemplo da seleção dos meses de fevereiro e agosto. As colunas representam a irradiação horária média de cada ano, a linha horizontal contínua representa a irradiação média entre - anos (média dos valores médios dos 13 anos de dados) com o desvio padrão.

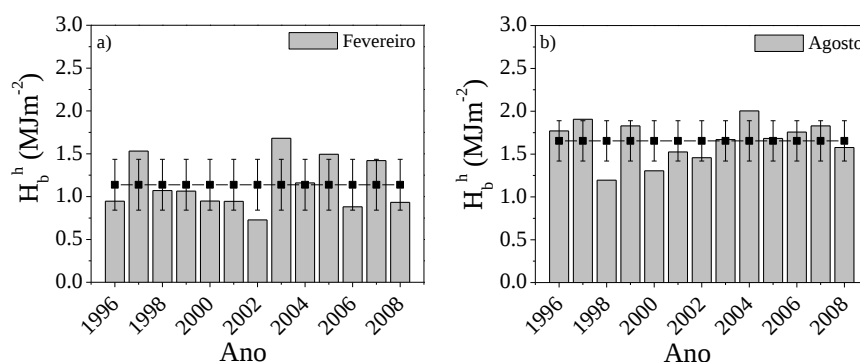


Figura 3. Comparação do valor da irradiação média entre - anos com desvio padrão e o valor da irradiação média do ano, para os meses: a) Fevereiro e b) Agosto.

O processo de seleção da base de dados do AT é similar ao processo de seleção do Ano Meteorológico Típico (WMO, 1981). Para o mês de fevereiro, o AT foi 2004, enquanto que em agosto foi 2003. Nestes dois meses, os valores das irradiações médias inter-anos praticamente coincidiram com os valores das irradiações médias do ano. Ao contrário, no AAT, os valores das irradiações médias do

ano estão mais distantes (para acima ou baixo) dos valores das irradiações médias inter-anos. Assim para o mês de fevereiro, o AAT foi 2003, enquanto que para agosto foi 1998. A constituição da base dos meses para o AT e AAT é representada na Tabela 3.

Tabela 3. Base que constituem o ano típico (AT) e ano atípico (AAT).

Meses												
	J	F	M	A	M	J	J	A	S	O	N	D
AT	04	04	98	06	07	98	00	03	96	97	06	04
ATT	98	03	00	08	03	03	08	98	03	03	98	96

5.3.6 Cálculo H_0 , do fotoperíodo (N) e H_{sc}^d

A irradiação global no topo da atmosfera (H_0 , MJm^{-2}) foi calculada de acordo com equações descritas em (SOUZA et al., 2005), Equação (1):

$$H_0 = I_0 \left[1 + 0,033 \cos \left(\frac{2\pi DJ}{365} \right) \right] \times [(\sin \varphi \sin \delta) + (\cos \varphi \cos \delta \sin \omega_s)] \quad (1)$$

em que I_0 é a constante solar ($= 1367 \text{ Wm}^{-2}$), DJ é o dia juliano iniciando em 1 de janeiro (para primeiro de janeiro, DJ = 1, e para 31 de dezembro, DJ = 365), φ é a latitude local (em graus), δ é a declinação solar (em graus) Eq. (2), ω_s é o ângulo solar horário do nascer do sol (em graus) Equação (3).

$$\delta = 23,45 \times \sin \left[\frac{360(DJ+284)}{365} \right] \quad (2)$$

$$\omega_s = \cos^{-1}[-\tan(\varphi) \times \tan(\delta)] \quad (3)$$

O fotoperíodo (N, em horas) é calculado conforme a Equação(4):

$$N = \frac{2}{15} \times \omega_s \quad (4)$$

A irradiação direta na incidência normal no topo da atmosfera diária (H_{sc}^d , MJm^{-2}) é obtida multiplicando a constante solar integrada ($H_{sc} = 4,921 \text{ MJm}^{-2}$) pelo fotoperíodo, (Eq. 5):

$$H_{sc}^d = 4,921 \times N \quad (5)$$

5.3.7 Indicadores de validação

O *Relative Mean Bias Error* (rMBE), *Relative Root Mean Square Error* (rRMSE), erro percentual relativo (e), *d* de Willmott e coeficiente de

determinação (R^2), foram utilizados para certificar a precisão dos modelos desenvolvidos (Escobedo et al., 2011; Santos et al., 2014). Valores positivos (negativos) de rMBE indicam superestimativa (subestimativa). O rRMSE dá a precisão a curto prazo das estimativas com relação aos dados medidos. Um erro percentual relativo (e) entre -10 e +10 % na validação de modelos é considerado aceitável (Ertekin e Yaldiz, 2000). O índice de ajustamento “d” varia de 0 a 1 e representa a total discordância ou a concordância perfeita, respectivamente, entre as medidas e as estimativas (WILLMOTT, 1982). Os índices são obtidos pelas relações:

$$rMBE = 100 \times \frac{\frac{\sum_{i=1}^x (H_E - H_M)}{x}}{\bar{X}} \quad (6)$$

$$rRMSE = 100 \times \frac{\left[\frac{\sum_{i=1}^x (H_E - H_M)^2}{x} \right]^{\frac{1}{2}}}{\bar{X}} \quad (7)$$

$$e = \left[\frac{(\bar{H}_M - \bar{H}_E)}{\bar{H}_M} \right] \times 100 \quad (8)$$

$$d = 1 - \frac{\sum_{i=1}^x (H_E - H_M)^2}{\sum_{i=1}^x (|H'_E| + |H'_M|)^2} \quad (9)$$

em que: H_E representa os valores estimados, H_M os valores medidos, $|H'_E|$ o valor absoluto da diferença $H_E - H_M$, em que $\bar{H}_M$ representa a média de H_M , $|H'_M|$ representa o valor absoluto da diferença $H_M - \bar{H}_M$ e $\bar{H}_E$ a média dos valores estimados. $\bar{X} \left(= \frac{1}{x} \sum_{i=1}^x H_M \right)$ é o valor médio da medida e x o número de observações. Recomendações quantitativas são por vezes usadas para validação de modelos: rMBE dentro de $\pm 10\%$ e $rRMSE < 20\%$, que indica bom ajuste entre as estimativas por modelos e as medidas na superfície (Badescu et al., 2013). Neste trabalho, os seguintes critérios são utilizados para classificar os modelos (Jamieson et al., 1991; Heinemann et al., 2012):

Excelente se (excellent if) $rRMSE < 10\%$;

Bom se (good if) $10\% \leq rRMSE < 20\%$;

Aceitável se (fair if) $20\% \leq rRMSE < 30\%$;

Pobre se (poor if) $rRMSE \geq 30\%$.

O R^2 fornece uma medida da relação linear entre os valores medidos e estimados. Os valores de R^2 variam de 0 a 1, e quanto maior seu valor, melhor o ajuste do modelo:

$$R^2 = \frac{\sum_{i=1}^x (H_M - \bar{H}_M)^2 - \sum_{i=1}^x (H_E - \bar{H}_M)^2}{\sum_{i=1}^x (H_M - \bar{H}_M)^2} \quad (10)$$

5.4 RESULTADOS E DISCUSSÃO

Neste trabalho, o algoritmo SMO foi utilizado para estimar a fração transmitida da irradiação direta na incidência normal (kt_b). O desempenho do SMO foi comparado com modelos estatísticos. Os resultados obtidos são desenvolvidos e discutidos, primeiramente, para a dispersão dos valores estimados com base na validação do AT e AAT, em seguida são analisados os valores de rMBE, rRMSE e d. Os resultados mais significativos são mostrados.

Primeiramente foram encontrados os valores dos parâmetros da função RBF de Kernel. Os parâmetros da RBF: C , ε e γ foram ajustados através de inúmeros ensaios com diferentes combinações. Foram considerados os seguintes conjuntos de valores dos parâmetros C , ε e γ : $C = \{10, 50, 100, 200 \text{ e } 400\}$, $\varepsilon = \{0,001; 0,002; 0,003; 0,004 \text{ e } 0,005\}$ e $\gamma = \{0,1; 0,2; 0,3; 0,4 \text{ e } 0,5\}$. Portanto, para SMO com a função RBF de Kernel foram realizados 60 simulações, empiricamente abrangendo as diferentes combinações dos parâmetros. Os valores com melhor ajuste nos modelos foram utilizados: $C = 50$, $\varepsilon = 0,005$ e $\gamma = 0,2$. A Tabela 4 mostra o R^2 de ajuste dos modelos de SMO. Os maiores valores de R^2 indicam os modelos gerados estão bem correlacionados. Os modelos apresentam elevados coeficientes de determinação ($R^2 > 0,867$).

Tabela 4. Coeficiente de determinação dos modelos SMO.

Modelo	SMO1	SMO2	SMO3
R^2	0,88	0,89	0,87

5.4.1 Análise de precisão dos modelos SMO1, SMO2 e SMO3

A dispersão de kt_b^h medido e estimado, com os modelos SMO1 e SMO2 são lineares com a linha ideal de comparação (1:1) (Figuras 4 a, c), nas duas bases de validação (AT e AAT). Os resultados com o ano típico serão identificados com (*) e os resultados com o ano atípico serão identificados com (**). O modelo SMO1 estimou kt_b^d com valores entre 0,05 e 0,626. Isto mostra que o algoritmo SMO gera boa estimativa de kt_b^h . Já as distribuições de frequência dos resíduos (Δ) são mostradas na Figura (4 b, d). O Δ é a diferença entre as medidas e as estimativas. Os valores de Δ para estes modelos tem frequências para valores próximos de zero, com uma

distribuição simétrica. São observados maiores valores de Δ com SMO1, isto pode estar relacionado com as condições atmosféricas de nebulosidade em que kt está sob condição de céu parcialmente nublado e/ou tendendo a limpa. Nos resultados da dispersão horária, já observados que valores baixos da massa ótica (m_a) ($1 - 1,25$) o kt_b^h pode chegar a 0,75, para valores médios de m_a ($1,25 - 2,5$) o kt_b^h chega a 0,70 e valores altos de m_a ($2,50 - 5,58$) o kt_b^h não supera 0,60 (YOUNES et al., 2005). De fato, nos valores elevados de m_a , as medidas e/ou estimativas de kt_b^h tendem a zero devido ao efeito de espalhamento e absorção na atmosfera.

Os aerossóis são, depois das nuvens, os principais atenuadores de H_b^h (kt_b^h) e responsáveis por essa redução (CEBECAUER et al., 2011). As análises horárias e diárias nas bases de AT e AAT podem ser úteis em estudos de impacto na utilização da tecnologia fotovoltaica e de concentradores solares. Pois, esta abordagem considera duas condições adversas dos valores esperados da radiação.

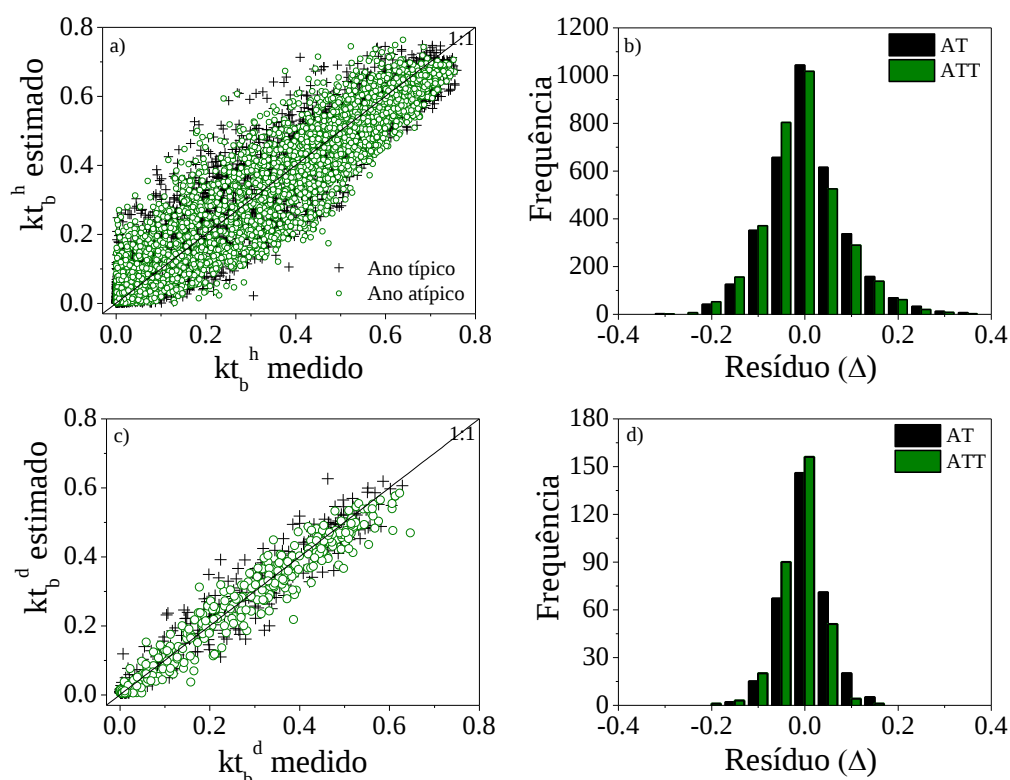


Figura 4. Dispersão dos valores de kt_b medido versus estimado através de: a) SMO1 e c) SMO2. Distribuição de freqüência dos Resíduos (Δ): b) SMO1 e d) SMO2. Ano típico (AT) e Ano Atípico (AAT).

O *SMO1 superestima em 1,79% e **SMO1 subestima em -1,51% (Figura 5a). A validação do modelo horário teve maior espalhamento com a base de dados do AT (*SMO1) $rRMSE = 26,28\%$, já o AAT (**SMO1) mostra $rRMSE = 24,23\%$ (Figura 5b). O índice “d” foi elevado e igual nas duas bases de validação $d = 0,97$ (Figura 5c). O espalhamento é explicado pela partição de tempo considerado, quantidade de dados utilizados na validação e pela variação temporal de H_b^h e H_G^h causada pelas nuvens e pelos aerossóis. Assim, valores de H_b em partições de tempo menores são mais difíceis de gerar modelos, por serem mais sensível à variabilidade temporal nas diferentes condições atmosféricas (Gueymard, 2010). Portanto, sendo uma estimativa horária, o SMO1 estima com aceitável precisão.

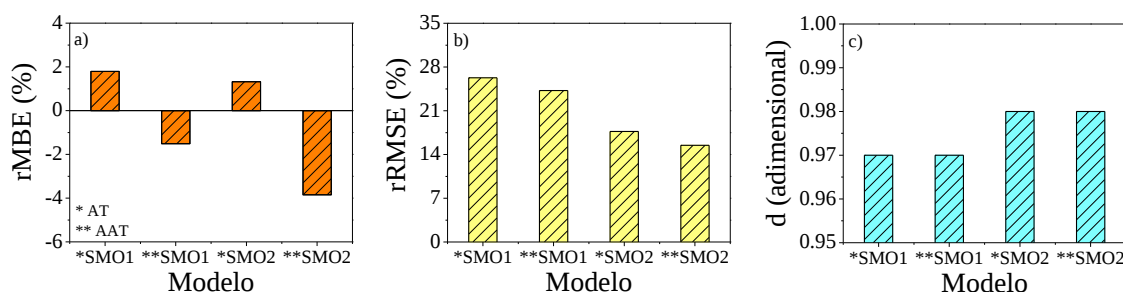


Figura 5. Índices de validação dos SMO1 e SMO2. *resultados com ano típico (AT). **resultados com ano atípico (AAT).

O modelo SMO2 tem boa generalização na estimativa de kt_b^d . Para *SMO2 os índices de validação $rMBE$, $rRMSE$ e d são 1,32%, 17,70% e 0,98, respectivamente. O **SMO2 subestimou as medidas em -3,85%, tem espalhamento de 15,50% e $d = 0,98$. O desempenho do modelo SMO2 melhor que SMO1 comprova a superioridade do modelo gerado na partição diária sobre a horária.

Na validação com AT, *SMO2 superestimou as medidas e com AAT **SMO2, subestimou. Neste caso, a superestimativa é causada pela maioria dos meses utilizados na elaboração do AT possuir valores médios de H_b abaixo da média de cada mês e, na base do AAT, a maioria dos meses tiveram valores médios acima da média de cada mês, ocasionando a subestimativa. Considerando o índice $rRMSE$, os resultados obtidos mostram que SMO2 estima com boa precisão ($rRMSE < 20\%$).

Os valores dos índices de validação ($rMBE$, $rRMSE$ e d) e o gráfico de dispersão do modelo SMO3, validado com AT e ATT, são apresentados na Figura (6a). A estimativa com o SMO3 mostra que os pontos crescem com a linha ideal

de comparação. A distribuição de frequência de Δ é simétrica com a maior valores próxima de zero (Figura 6b).

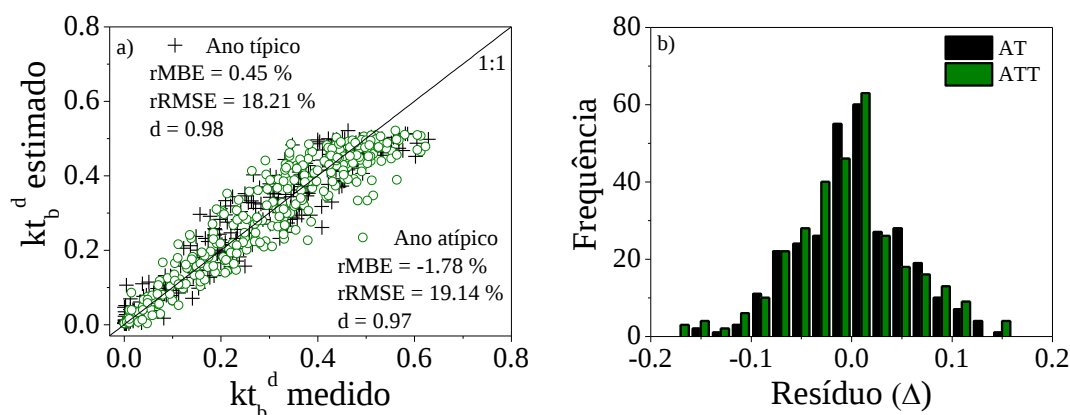


Figura 6. Dispersão e índices de validação do modelo SMO3 na estimativa de kt_b^d (a) e Δ (b).

Considerando o rRMSE, o SMO3 estima kt_b^d com boa precisão. A validação com o AT (*SMO3) estima com: rMBE = 0,45%, rRMSE = 18,21% e d = 0,98 e **SMO3 com (rMBE = -1,78%, rRMSE = 19,14% e d = 0,97) (Figura 6a). Similar à validação dos modelos SMO1 e SMO2, o modelo SMO3 com o AAT subestimou as medidas e com AT superestimou. Na estimativa diária, o desempenho do SMO com a relação kt_b^d vs kt é melhor que o desempenho com a relação kt_b^d em função de n/N . Isto evidencia que kt é uma variável importante na modelagem de kt_b . Assim, o mais provável é que problemas na queima do heliograma e erros operacionais na leitura afete a estimativa de kt_b^d a partir de n/N .

5.4.2 Análise de precisão dos modelos

O gráfico de dispersão mostra alta correlação entre os valores estimados de kt_b por #M1 (Figura 7a) e #M2 (Figura 7c). A maior dispersão com #M1 ocorre no intervalo $(0,20 < k_b^h < 0,50)$, onde se encontra uma maior densidade de ponto. É notório a influencia das condições de cobertura de céu sobre os modelos, quando se observa valores baixos, médios e elevados. Por isso, os valores de H_b em partição de tempo menor são mais difíceis de modelar devido à variabilidade da cobertura de nuvens e de outros constituintes atmosféricos (Padovan e Del Col, 2010). A distribuição de frequência para os valores de Δ para #M1 é simétrico (Figura 7b). Nos valores baixos

de kt_b^h , a distribuição de frequência de Δ mostra que as estimativas são maiores que as medidas. Enquanto que, para os valores altos, nota-se que as estimativas são menores que as medidas. Isto é causado pelo efeito das coberturas de nuvens. Já que uma situação corresponde ao período de céu nublado e na outra a condição de céu claro.

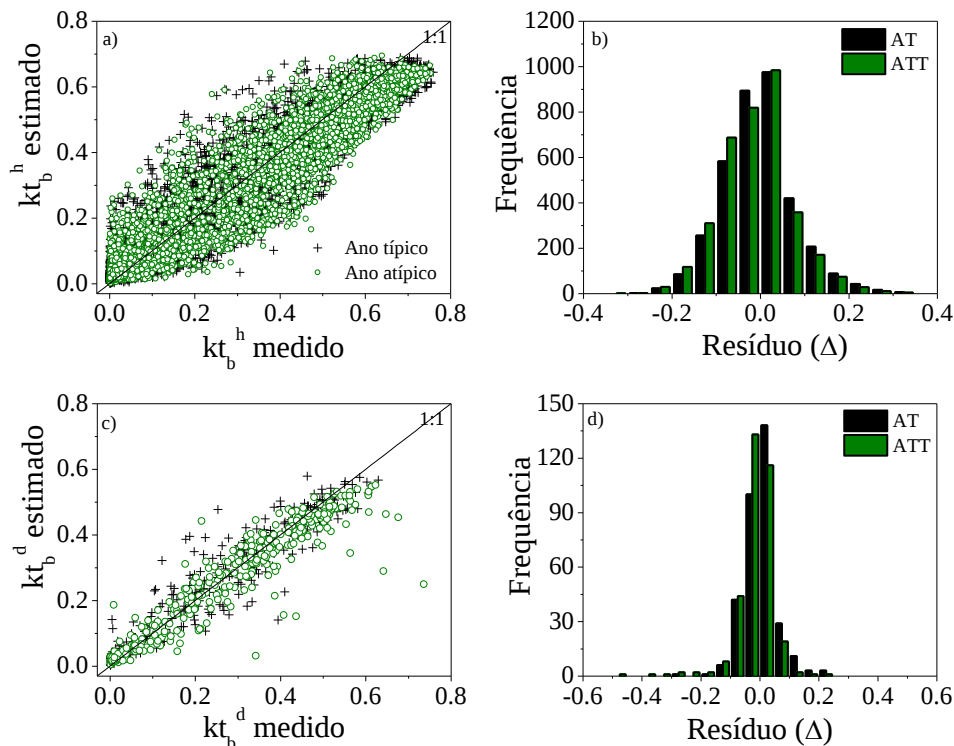


Figura 7. Dispersão de kt_b medido e estimado, com a frequência de Δ , respectivamente, para o modelo #M1 (a) e (b) frequência de Δ , modelo #M2 (c) e (d) frequência de Δ .

A estimativa com o #M2 apresenta alta correlação. A distribuição de frequência de Δ tem maiores valores próximos de zero (Figura 7d). A dispersão é explicada pelas nuvens e pelos processos físicos de transferência radiativa na atmosfera (espalhamento e absorção pelos constituintes atmosféricos), pois existe um grande número de combinações de cobertura de nuvens e aerossóis que irá gerar o mesmo valor de kt , mas não se pode esperar que cada combinação produza um valor único de kt_b (BALARAS, 1990).

Os valores dos índices rMBE, rRMSE e d dos modelos #M1 e #M2, nas bases de AT e AAT, são apresentados nas Figuras 8 (a, b, c). Os dois modelos têm desempenho aceitáveis ($20\% \leq rRMSE < 30\%$). O *#M1 possui maior espalhamento ($rRMSE = 26,51\%$), menor subestimativa ($rMBE = -0,61\%$) e índice de concordância $d = 0,97$. Similarmente aos modelos de SMO, esperava-se que *#M1

resultasse em superestimativa das medidas. Contudo, esta subestimativa pode estar relacionada com o baixo desempenho do modelo estatístico quando comparado com o algoritmo SMO. A validação com ****#M1** resultou em rMBE e d iguais a -3,73% e 0,97, respectivamente. Utilizando o modelo de Louche et al., (1991), Bertrand et al., (2015) obtiveram na Belgica valores de kt_b com erros elevados (rMBE = 14,90% e rRMSE = 31,90%). Estes são maiores que os valores obtidos neste trabalho. Outros modelos avaliados por Bertrand et al., (2015) também apresentaram erros maiores aos obtidos com **#M1** e aos modelos de SMO para Botucatu. A validação com ****#M2** tem a maior subestimativa (rMBE = -6,07%) e d = 0,96.

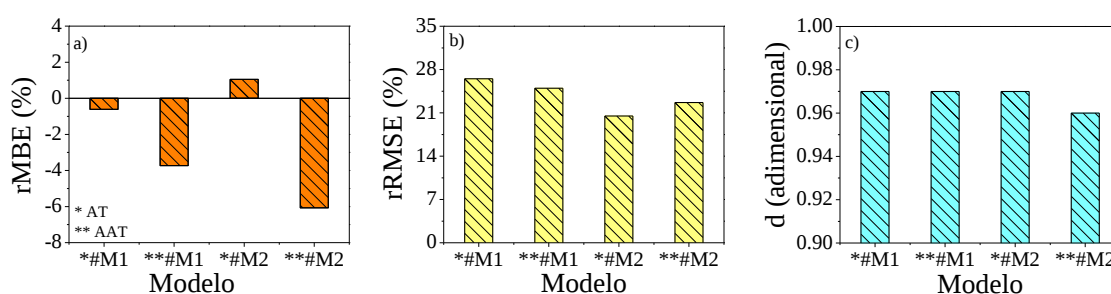


Figura 8. Índices de validação dos modelos **#M1** e **#M2**. (*) Ano típico e (**) Ano Atípico.

A dispersão e os valores dos índices de validação do modelo **#M3** são mostrados na Figura (9a). A dispersão de kt_b^d medido em função da estimativa é maior para valores de $kt_b^d > 0,4$. A distribuição de frequência do resíduo (Δ) de **#ME3** é simétrica (Figura 9b). A validação com ****#M3** apresentou rRMSE = 22,89% e subestimativa (rMBE = -6,86%). O ***#M3** subestimou (rMBE = -3,45%), rRMSE = 20,79% e d = 0,96. Com base nos valores de rRMSE, o desempenho de **#M3** estima com precisão aceitável. Os resultados aqui apresentados são satisfatórios e o gráfico de dispersão mostra a limitação que o **#M3** possui quando comparado com a SMO3.

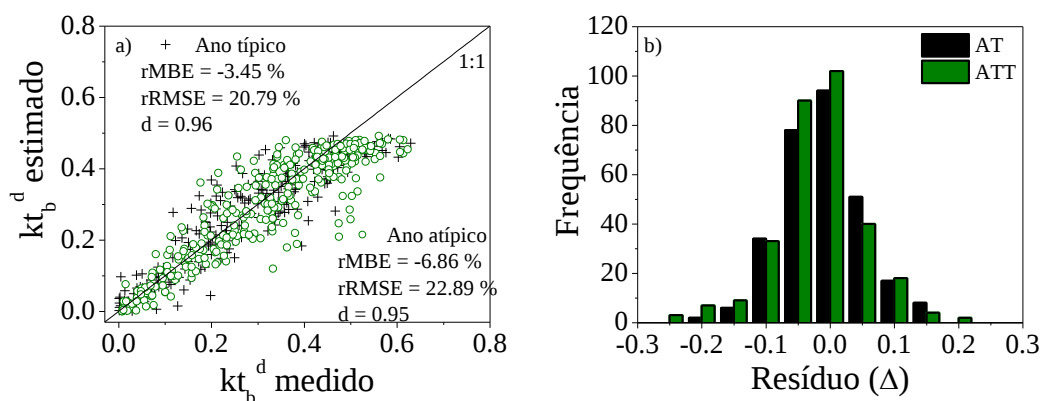


Figura 9. Dispersão e valores dos índices de validação do modelo #M3 na estimativa de kt_b^d , (a) dispersão e (b) Δ .

5.4.3 Análise comparativa entre a precisão de SMO e os modelos estatísticos

Para verificar a adequação da SMO na estimativa de kt_b , as precisões comparam-se com os modelos estatísticos. Esta comparação é feita para mostrar qual a metodologia tem melhor estimativa de kt_b .

Os resultados de $rMBE$ obtidos com SMO já discutidos, em média, são menores que dos modelos estatísticos. Nas duas bases de validações com SMO1 – SMO3 os valores de $rRMSE$ são (0,28 – 4,99%) menores que os modelos estatísticos (#M1, #M2 e #M3). O índice “d” obtido com SMO discutidos anteriormente é semelhante aos obtidos com os modelos estatísticos. A validação dos modelos da SMO na base do AT tem predominância de superestimar e com o AAT em subestimar.

A Figura 10 mostra o erro percentual relativo (e) obtido na validação dos modelos. Considerando os limites de variação de e ($\pm 10\%$), os resultados obtidos são aceitáveis para os modelos aqui avaliados. Em geral, a validação com os modelos de SMO mostram menores valores de e quando comparado com os resultados dos modelos #M. Nos modelos de SMO validados com AT os valores de e foram negativos e quando validados com AAT foram positivos. Maior desempenho dos modelos gerados com as técnicas de AM, com relação aos modelos empíricos (estatísticos) são destacados destacada na literatura (Mohammadi et al., 2015; Chen et al., 2015; Shamshirband et al., 2015). Por fim, a comparação pareada dos modelos aqui apresentados mostra que SMO tem potencial para estimar kt_b .

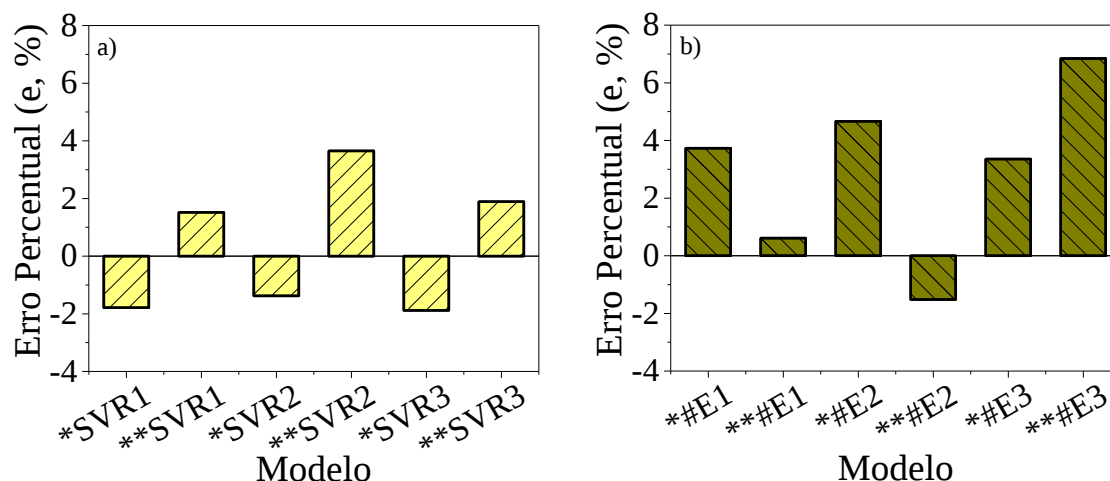


Figura 10. Erro percentual relativo (e, %) dos modelos: a) SMO e b) #M. (*) Ano típico e (**) ano atípico.

5.4.4 Análise de precisão de SMO nas estações chuvosa e seca: estudo de caso.

Nesta seção um estudo de caso para avaliação de SMO é desenvolvido. A série de 13 anos é separada em duas: estação chuvosa e estação seca. Do total dos dados utilizados de cada estação, 80% foram utilizados para treinamento e 20% na validação. Esta divisão é considerada para analisar o efeito das nuvens, dos aerossóis e do vapor d'água no desempenho da SMO e na dispersão de kt_b . Este estudo é motivado pela região de Botucatu ser fortemente afetada pela queima de cana de açúcar para a colheita mecanizada e queimadas florestais, que aumentam significativamente a concentração de materiais particulados na atmosfera local (Codato et al., 2008). O vapor d'água na atmosfera tem origem das represas hidroelétricas situadas num raio de 70 km de Botucatu e, muito provavelmente, das usinas sucroalcooleira no período da colheita e moagem em regiões adjacentes. Nesta seção, os modelos são simbolizados como SMO1.1 (kt_b^h em função de kt, horário), SMO2.2 (kt_b^d em função de kt, diário) e SMO3.3 (kt_b^d em função de n/N, diário).

As Figuras 10 (a, b, c) mostram os valores dos índices de validação e a dispersão das estimativas para os modelos na estação chuvosa e na estação seca. Na estação chuvosa SMO1.1 gerou maiores erros ($rMBE = 2,45\%$, $rRMSE = 30,89\%$ e $d = 0,96$) que na estação seca ($rMBE = 1,87\%$, $rRMSE = 21,80\%$ e $d = 0,97$) (Figura 11a). Neste modelo, observa-se que os pontos são mais dispersos na estação chuvosa que na seca. Assim, o resultado confirma que na estação chuvosa tem aumento

de $\approx 9\%$ do espalhamento de kt_b^h com relação à estação seca. Este resultado é devido ao efeito das nuvens, principalmente. Eliminado as nuvens, os aerossóis de queimadas agrícolas e florestais, assim como, o vapor d'água das represas podem ser os fatores responsáveis pelo espalhamento da kt_b^h . Analisando os valores de rRMSE, na estação chuvosa a estimativa tem precisão pobre e na estação seca precisão aceitável.

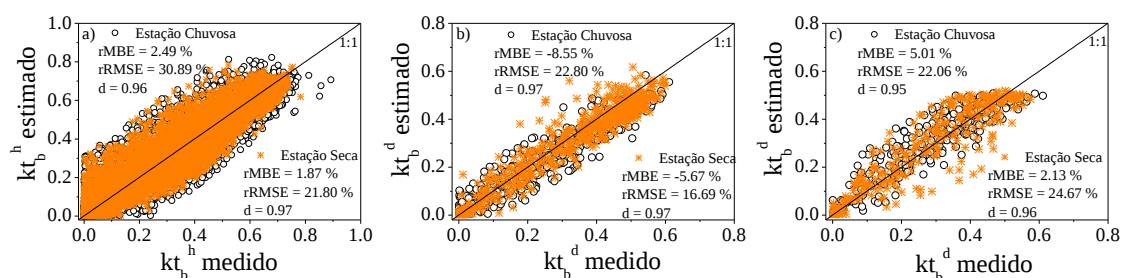


Figura 11. Dispersão e índices de validação. a) SMO1.1, b) SMO2.2 e c) SMO3.3.

Na estimativa com SMO2.2, os índices de validação na estação chuvosa têm $rMBE = -8,55\%$, $rRMSE = 22,80 \%$ e $d = 0,97$ e na estação seca: $rMBE = -5,67$, $rRMSE = 16,69 \%$ e $d = 0,97$ (Figura 11b). Neste modelo, a dispersão tem o melhor ajuste da estimativa para a estação seca. Estes resultados mostram o efeito que as nuvens tem no desempenho de SMO2.2 na estação chuvosa, principalmente, nos valores de kt_b^d menores que 0,40. Na estação chuvosa a estimativa é aceitável e na estação seca o modelo estima com boa precisão. O SMO3.3 estima com aceitáveis rRMSE nas duas estações (Figura 11c). O valor do rRMSE na estação chuvosa (22,06%) é menor que na estação seca (24,67%). Os resultados observados com SMO2.2 e SMO3.3 são justificados pelo aumento das nuvens no período chuvoso, enquanto que, na ausência das nuvens, os aerossóis e o vapor d'água elevam o espalhamento. Com relação à variável de entrada, kt estima com erros menores que n/N . Por fim, o SMO mostra ser eficiente para estimativa de kt_b nas diferentes condições atmosféricas e tem potencial para aplicação em locais de interesse com resultados satisfatórios.

5.5 CONCLUSÕES

O trabalho comparou o desempenho dos modelos gerados pela AM SVM, com uso do algoritmo SMO com modelos estatísticos nas partições horária e

diária. Os modelos SMO1, SMO2 e SMO3 apresentaram bons ajustes na modelagem, com elevados coeficientes de determinação (R^2). Os valores de R^2 gerados pelos modelos de SMO são maiores que dos modelos estatísticos. Para a função de Kernel RBF os parâmetros utilizados e recomendados na modelagem de kt_b são: $C = 50$, $\gamma = 0,2$ e $\varepsilon = 0,005$. As variáveis de entrada foram capazes de modelar kt_b . Na estação seca os aerossóis e o vapor d'água são responsáveis pela atenuação da radiação e no aumento de rRMSE. Já na estação chuvosa, o aumento de 9% no espalhamento é causado pela atenuação da radiação pelas nuvens. A comparação mostra desempenho superior dos modelos de SMO com relação aos estatísticos. O SMO é indicado para estimativa da kt_b como a primeira opção devida sua capacidade de generalização e adequação.

Os resultados mostraram que o SMO pode servir como uma alternativa promissora para resolver problemas com programação quadrática para estimar kt_b , com capacidade de fornecer maior precisão em comparação com outras técnicas existentes. O SMO também pode ser utilizado para a estimativa de kt_b em regiões para as quais não existem dados de radiação e tenha condições climáticas semelhantes a Botucatu-SP. Os resultados obtidos, também, agregam importância da aplicação de SMO para recuperação de série histórica. A geração de uma série histórica vai permitir um estudo mais detalhado da variabilidade climática na região considerada, assim como, no desenvolvimento de projetos e no planejamento de centrais solares para geração de energia.

Estudos futuros são considerados para avaliação dos diferentes algoritmos de aprendizado de máquinas existentes no WEKA para estimar de outras componentes da radiação solar. Outras variáveis de entrada na modelagem de kt_b e H_b devem ser avaliadas para a ampliação da generalização da técnica e da irradiação. Avaliação futura é necessária para estimativa de H_b em diferentes escalas de tempo, principalmente em partições menores.

5.6 REFERÊNCIAS BIBLIOGRÁFICAS

- ÅNGSTRÖM, A. K. Solar and terrestrial radiation. **Quarterly Journal of the Royal Meteorological Society**, v.50, p.121, 1924.
- BADESCU, V.; GUEYMARD, C. A.; CHEVAL, C.; OPREA, C.; BACIU, M.; DUMITRESCU, A.; IACOBESCU, F.; MILOS, I.; RADA, C. Accuracy analysis

forty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. **Renewable Energy**, v.55, p.85-103, 2013.

BALARAS, C. A. Realistic estimates of the beam radiation from global radiation measurements through an enhanced model for Athens, Greece. **Solar & Wind Technology**, v. 7, p.313–320, 1990.

BEHAR, O.; KHELLAF, A.; MOHAMMEDI, K. Comparison of solar radiation models and their validation under Algerian climate – The case of direct irradiance. **Energy Conversion and Management**, v. 98, p.236-251, 2015.

BERTRAND, C.; VANDERVEKEN, G.; JOURNEE, M. Evaluation of decomposition models of various complexity to estimate the direct solar irradiance over Belgium. **Renewable Energy**, v.74, p.618-626, 2015.

BEYER, H.G.; FAUTER, M.; SCHUMANN, K.; SCHENK, H.; MEYER, R. Synthesis of DNI time series with sub-hourly time resolution. **In: 16th SolarPACES Conference. Perpignan (France)**, 2010.

BLANC, P. ET AL. Direct normal irradiance related definitions and applications: the circumsolar issue. **Solar Energy**, v.110, p.561–77, 2014.

BURGES, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. **Data Mining and Knowledge Discovery**, v.2, p.121-167, 1998.

CEBECAUER, T, SÛRI M, GUEYMARD, A. C. Uncertainty sources in satellite-derived direct normal irradiance: how can prediction accuracy be improved globally?. **In: Proceedings of the Solar PACES 2011 Conference: Granada, Spain; September 2011.**

CHEN, J-L.; LI, G-S.; XIAO, B-B.; WEN, Z-F.; LV, M-Q.; CHEN, C-D.; JIANG, Y.; WANG, X-X.; WU, S-J. Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. **Energy Conversion and Management**, v.89, p. 318–329, 2015.

CHEN, J-L; LI, G-S, WU, S-J. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. **Energy Conversion and Management**, v.75, p. 311–318, 2013.

CHEN, J-L; LI, G-S. Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. **Theoretical and Applied Climatology**, v.115, p.627–638, 2014.

- CODATO, G.; OLIVEIRA, A. P.; SOARES, J.; ESCOBEDO, J. F.; GOMES, E. G.; DAL PAI, A. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. **Theoretical and Applied Climatology**, v. 93, p.57–73, 2008.
- ERTEKIN, C.; YALDIZ, O. Comparison of some existing models for estimating global solar radiation for Antalya (Turkey). **Energy Conversion and Management**, v. 41, p. 311-330, 2000.
- ESCOBEDO, J. F.; GOMES, E. N.; OLIVEIRA, A. P.; SOARES, J. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. **Renewable Energy**, v.36, p.169-178, 2011.
- FERNANDEZ-PERUCHEMA, C. M.; BLANCO, M.; GASTON, M.; BERNARDOS, A. Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones. **Solar Energy**, v. 115, p. 225-263, 2015.
- FERNANDEZ-PERUCHENA, C.; BLANCO, M.; BERNARDOS, A. Generation of series of high frequency DNI years consistent with annual and monthly long-term averages using measured DNI data. **Energy Procedia**, v.49, p. 2321–2329, 2014.
- GRANTHAM, A. P.; PUDNEY, P. J.; BOLAND, J. W.; BELUSKO, M. Synthetically interpolated five-minute direct normal irradiance. In: **20th International Congress on Modelling and Simulation**. Adelaide, Australia, 2013.
- GUEYMARD C, A. Progress in direct irradiance modeling and validation. In: **Proceedings of Solar 2010 Conference, Phoenix, AZ**; American Solar Energy Society; 2010.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN. I. H. **The WEKA Data Mining Software: An Update**; SIGKDD Explorations, Volume 11, Issue 1, 2009.
- HEINEMANN, A. B.; VAN OORT, P. A. J.; FERNANDES, D. S.; MAIA, A. H. N. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. **Bragantia**, Campinas, v. 71, n. 4, p.572-582, 2012.
- JAMIESON, P. D.; PORTER, J. R.; WILSON, D. R. A test of the computer simulation model ARC-WHEAT1 on wheat crops grown in New Zealand. **Field Crops Research**, v.27, p.337-350, 1991.
- JOURNÉE, M.; BERTRAND, C. Quality control of solar radiation data within the RMIB solar measurements network. **Solar Energy**, v.85, p. 72-86, 2011.

- KOTTI, M. C.; ARGIRIOU, A. A.; KAZANTZIDIS, A. Estimation of direct normal irradiance from measured global and correct diffuse horizontal irradiance. **Energy**, v.70, p.382-392, 2014.
- LIU, B. Y. H.; JORDAN, R. C. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. **Solar Energy**, v.3, p.1-19, 1960.
- LORENA, A. C.; JACINTHO, L. F. O.; SIQUEIRA, M. F.; GIOVANNI, R.; LOHMANN, L. G.; CARVALHO, A. C. P. L. F.; YAMAMOTO, M. Comparing machine learning classifiers in potential distribution modelling. **Expert Systems with Applications**, v.38, p.5268–5275, 2011.
- LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma Introdução às Support Vector Machines. **RITA**, v. 14, n. 2, p. 43-67, 2007.
- LOUCHE, A.; NOTTON, G.; POGGI, P.; SIMONNOT, G. Correlations for direct normal and global horizontal irradiation on french mediterranean site. **Solar Energy**, v.46, p.261-266, 1991.
- MAXWELL E, L. **A quasi-physical model for converting hourly global to direct normal insolation**. SERI/TR-215-3087, Solar Energy Research Institute, Golden, CO; 1987.
- MOHAMMADI, K.; SHAMSHIRBAND, S.; ANISI, M. H.; ALAM, K. A.; PETKOVIC, D. Support vector regression based prediction of global solar radiation on a horizontal surface. **Energy Conversion and Management**, v.91, p.433–441, 2015.
- OLIVEIRA, L. L. A. Estimation of software project effort with Support Vector Regression. *Neurocomputing*, v. 69, p. 1749-1753, 2006.
- PADOVAN, A.; DEL COL, D. Measurement and modeling of solar irradiance components on the horizontal and tilted planes. **Solar Energy**, v.84, p. 2068-2084, 2010.
- PADOVAN, A.; DEL COL.; SABATELLI, V.; MARANO, D. DNI estimation procedures for the assessment of solar radiation availability in concentrating systems. **Energy Procedia**, v. 57, p.1140-1149, 2014.
- PEREZ, R.; SEALS, R.; ZELENKA, A.; INEICHEN, P. Climatic evaluation of models that predict hourly direct irradiance from hourly global irradiance: prospects for performance improvements. **Solar Energy**, v.44, p.99-108, 1990.
- PIRI, J.; SHAMSHIRBAND, S.; PETKOVIC, D.; TONG, C. W.; REHMANB, M. H. Prediction of the solar radiation on the Earth using support vector regression technique. **Infrared Physics & Technology**, v.68, p.179–185, 2015.

- PLATT, J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. **Redmond: Microsoft Research**, 1998. 21 p. TechReport.
- PRESCOTT, J. A. Evaporation from a water surface in relation to solar radiation. **Transaction of the Royal Society of Science Australia**, v.64, p.114–125, 1940.
- RAGHAVENDRA, S.; DEKA, P. C. Support vector machine applications in the field of hydrology: A review. **Applied Soft Computing**, v.19, p.372–386, 2014.
- SANTOS, C. M. ET AL. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. **Energy**, v.71, p.388-398, 2014.
- SHAMSHIRBAND, S.; MOHAMMADI, K.; YEE, P. L.; PETKOVIĆ, D.; MOSTAFAEIPOUR, A. A comparative evaluation for identifying the suitability of extreme learning machine to predict horizontal global solar radiation. **Renewable and Sustainable Energy Reviews**, v.52, p.1031–1042, 2015.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v.14, p.199-222, 2004.
- SOUZA, J. L.; NICÁCIO, R. M.; MOURA, M. A. L. Global solar radiation measurements in Maceió, Brazil. **Renewable Energy**, v. 30, p. 1203 – 1220, 2005.
- VAPNIK, V. N. **Statistical learning theory**. New York: Wiley, 1998.
- VAPNIK, V. N. **The nature of statistical learning theory**. New York: Springer, 1995.
- VIANA, T. S.; RUTHERA, R.; MARTINS, F. R.; PEREIRA, E. B. Assessing the potential of concentrating solar photovoltaic generation in Brazil with satellite-derived direct normal irradiation. **Solar Energy**, v.85, p.486–495, 2011.
- WILLMOTT, C. J. Some comments on the evaluation of model performance. **Bulletin American Meteorological Society**, Boston, v.63, n.11, p.1309-1313, 1982.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. Data Mining: Practical Machine Learning Tools and Techniques. 3rd ed. 2011, 630p.
- WMO - World Meteorological Organization. Meteorological Aspects of the Utilization of Solar Radiation as an Energy Source. **World Meteorological Organization Technical Note No. 172**, WMO-No. 557, Geneva, p. 298, 1981.
- YADAV, A. K.; CHANDEL, S. S. Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model. **Renewable Energy**, v.75, p.675-693, 2015.
- YOUNES, S.; CLAYWELL, R.; MUNEER, T. Quality control of solar radiation data: Present status and proposed new approaches. **Energy**, v.30, p.1533–1549, 2005.

ZHANG, W. G.; ZHANG, X. L.; XIAO, W. L. Portfolio selection under possibilístico mean-variance utility and a SMO algorithm. **European Journal of operational Research**, v. 197, p. 693-700, 2009.

ZHANG, X.; ZHANG, W. G; XU, W. J. An optimization model the portfolio adjusting problem with fuzzy return and a SMO algorithm. **Expert Systems with applications**, v. 38, p. 3069 – 3074, 2011.

6. ESTIMATIVA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL: REDES NEURAIS ARTIFICIAIS E MODELOS ESTATÍSTICOS.

6.1 RESUMO

A irradiação direta na incidência normal (H_b) foi estimada utilizando modelos gerados Redes Neurais Artificiais (RNA) para uma base de dados medida do período de 1996 a 2008 em Botucatu- Brasil, região com influência de queimadas agrícolas. A rede utilizada foi a *Multilayer Perceptron* (MLP), com o algoritmo de aprendizado *backpropagation*. As RNAs foram comparadas com três modelos estatísticos de ajuste local. As variáveis de entrada foram o k_t (transmissividade atmosférica da irradiação global) e n/N (razão de insolação). Os modelos foram avaliados por meio dos indicadores estatísticos: *Mean Bias Error* (MBE, MJm^{-2}), *Relative Mean Bias Error* (rMBE, %), *Root Mean Square Error* (RMSE, MJm^{-2}), *Relative Root Mean Square Error* (rRMSE, %) e d de Willmott. A estimativa diária de H_b apresentou erros menores que as estimativas horárias. Os modelos que foram estimados com k_t como variável de entrada possui desempenho superior aos modelos de n/N . O desempenho das RNAs é melhor que dos modelos estatísticos. As estimativas com RNAs tiveram em média: $\text{rRMSE} = 19,59 \pm 2,87\%$ e $d = 0,95 \pm 0,02$. Os modelos estatísticos ajustados localmente estimam com erros médios: $\text{rRMSE} = 20,46 \pm 1,38\%$ e $d = 0,95 \pm 0,02$. Realizou-se um estudo de caso para condição de céu sem nuvens. Com os modelos das redes neurais, as

nuvens são responsáveis no aumento do rRMSE em $7,31 \pm 3,14\%$. As RNAs são potencialmente recomendadas como a primeira escolha para estimativa de H_b .

Palavras-Chave: WEKA, Mineração de Dados (Data Mining), Aprendizado de Máquina, Modelagem.

6.2 INTRODUÇÃO

A radiação solar é a principal fonte de energia renovável na biosfera, para os processos químicos, físicos e biológicos da Terra, importantes para o equilíbrio energético no sistema terra-atmosfera, na produção agropecuária, na disponibilidade de recursos hídricos e na modelagem (MANZANO et al., 2015; WANG et al., 2015). A irradiação direta na incidência normal (H_b) é destaque nos sistemas de conversão de energia solar devido à demanda energética mundial e a necessidade de novas fontes alternativas de energia renovável. A conversão da energia solar é feita através de células fotovoltaicas e dos sistemas de concentradores. Nos sistemas concentrados a H_b é destaque por ser a única componente com potencial para ser concentrada. Na conversão solar em térmica, a H_b , é aplicada no aquecimento e iluminação de ambientes, nos processos de geração de calor e no aquecimento de água (SANTOS et al., 2014a).

Valores de H_b são importantes em muitas áreas de pesquisas. Com a demanda mundial de dados solarimétricos, a modelagem tem sido utilizada com bastante frequência para estimativa de H_b . Estimativas de H_b dependem da variabilidade temporal e espacial. Por isso, é recomendado selecionar adequadamente o modelo para aplicação no local de interesse.

Na literatura existem muitos modelos estatísticos para estimativa de H_b (VIGNOLA:MCDANIELS, 1986; LOUCHE et al., 1991; NFAOUI:BURET, 1993; LÓPEZ et al., 2000). Estes modelos são baseados nas metodologias de Liu e Jordan (1960) e Ångström (1924). Os modelos de transferência radiativa são parametrizados para cada local e necessitam de informações sobre aerossóis, vapor d'água, gases e ozônio (GUEYMARD, 2003a; GUEYMARD, 2003b). Estas informações não estão facilmente disponíveis em superfície. Portanto, a exploração de novos métodos tem sido bastante difundido nos últimos anos.

As Redes Neurais Artificiais (RNAs) têm sido aplicadas com sucesso para previsão de irradiação solar e sistemas solares (QAZI et al., 2015). Koca et al., (2011) aplicaram RNAs na estimativa da H_g em sete locais da região do mediterrâneo de Anatólia. Wana et al., (2008) utilizaram regressão e RNAs para investigar as correlações entre a radiação solar global diária e a insolação para locais de diferentes climas na China. Em estudos para a cidade de São Paulo, Soares et al., (2004) utilizaram RNAs na estimativa da irradiação difusa e compararam os resultados com modelos estatísticos. Segundo os autores os modelos de RNAs tiveram melhor desempenho e pode ser utilizada na estimativa solar com resultados satisfatórios. Kalogirou et al., (2014) empregaram RNAs na previsão de grandes sistemas solares. Os resultados de comparação foram satisfatórios e a RNAs pode ser usado com sucesso. Em oito locais da região Nordeste do Brasil, Lyra et al., (2015) utilizaram RNAs e modelos empíricos para estimativa da irradiação solar global. Apesar dos inúmeros trabalhos com RNAs na estimativa de H_g (YADAV:CHANDEL, 2014), escassos são os trabalhos na estimativa de H_b .

O objetivo deste trabalho é investigar a viabilidade da utilização de RNA na estimativa de H_b . A técnica neural foi comparada com modelos estatísticos nas partições de tempo horário e diário. A rede neural possui as mesmas variáveis de entrada dos modelos estatísticos, a transmissividade atmosférica da irradiação global (k_t) e a razão de insolação (n/N). O trabalho tem a importância de mostrar a viabilidade das RNAs na estimativa e no mapeamento de H_b nos locais onde não existem medidas. O estudo está organizado da seguinte forma. A descrição dos dados, local de estudo e a metodologia adotada é dada na primeira parte. Nos resultados é mostrado o desempenho dos modelos, destacando a metodologia com melhor desempenho. A rede é avaliada, no estudo de caso, na condição de céu sem nuvens para verificar o efeito dos aerossóis e vapor d'água na dispersão. Por último as principais conclusões são destacadas.

6.3 MATERIAL E MÉTODOS

6.3.1 Local de estudo e dados

Os dados utilizados foram medidos na Estação Radiométrica localizada na Faculdade de Ciências Agrônômicas da UNESP/Botucatu (latitude 22,85°S, longitude 48,45°W e altitude 786m) (Figura 1a). Foram utilizados 13 anos

(1996 a 2008) de dados irradiação direta na incidência normal (H_b), irradiação global (H_g) e insolação (n). A H_b foi obtida por pireliômetro da Eppley NIP acoplado a um rastreador solar ST3 da Eppley, a H_g por piranômetro Eppley PSP e a insolação (n) solar por heliógrafo convencionais do tipo Campbell-Stokes (Figura 1b, c, d). Os sensores são aferidos periodicamente e todas as medições são feitas seguindo o guia da Organização Meteorológica Mundial (WMO, 1983).

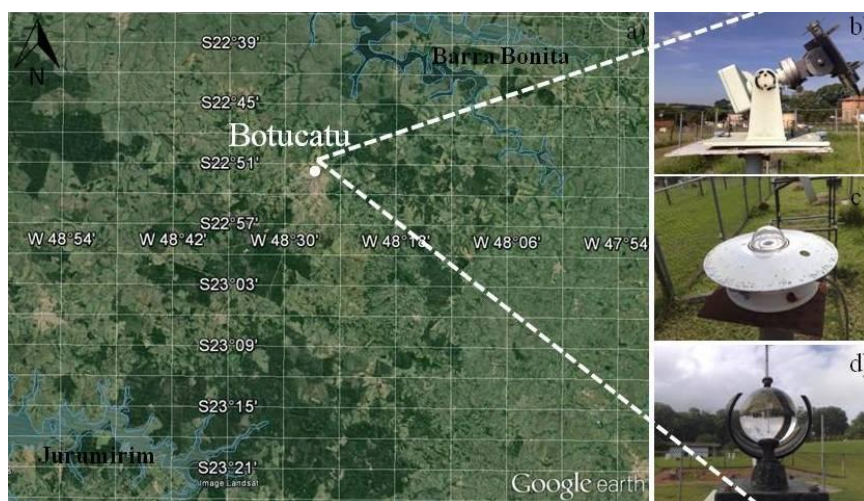


Figura 1. Local e dados utilizados. a) localização da estação, b) Pireliômetro, c) Piranômetro e d) Heliógrafo.

A região de Botucatu é caracterizada por possuir clima temperado quente (mesotérmico), verão quente e úmido com elevada precipitação e inverno seco (Escobedo et al., 2011). A Tabela 1 apresenta as séries climáticas da temperatura do ar, umidade relativa do ar e precipitação acumulada, do período de 1971 a 2009. O mês mais quente é fevereiro, o mais frio, julho, e com temperatura do ar média anual de $20,086 \pm 2,176^\circ\text{C}$. A umidade relativa do ar varia de $61,15 \pm 5,92\%$ (agosto) a $76,41 \pm 2,54\%$ (janeiro). A precipitação média anual acumulada é de 1381,29 mm. O período mais chuvoso ocorre de outubro a março (primavera-verão) e o menos chuvoso entre abril e setembro. No período chuvoso, a precipitação é causada pela Zona de Convergência do Atlântico Sul (ZCAS). No período seco, a precipitação é do tipo frontal originada do encontro das massas de ar frias e secas vindas da região sul com as massas quentes e úmidas, da região sudeste.

Tabela 1. Séries climáticas da temperatura (T_{ar}), umidade do ar (UR) e precipitação (P) no período de 1971 a 2009.

Mês	T_{ar} (°C)	UR (%)	P (mm)
Jan	22,13±0,81	76,41±2,54	305,75±120,03
Fev	22,53±0,72	75,15±1,38	188,38±79,54
Mar	22,19±0,55	72,81±2,51	143,38±61,25
Abr	20,96±1,03	70,48±2,23	58,82±34,98
Mai	17,22±0,99	67,90±2,00	75,54±45,95
Jun	17,05±0,65	66,73±2,38	38,88±37,80
Jul	16,77±0,89	63,13±2,94	41,33±48,08
Ago	18,80±1,12	61,15±5,92	38,83±33,45
Set	19,28±0,83	64,76±4,28	70,65±51,65
Out	20,71±1,07	68,48±2,84	109,46±52,11
Nov	21,23±0,81	69,64±2,68	113,60±61,51
Dez	22,17±0,74	71,37±3,33	196,67±44,40

Nos meses da estação seca foi observada a prática da queima de cana-de-açúcar em cidades adjacentes a Botucatu, que são realizadas para a colheita dos colmos (Codato et al., 2008). Com a queima de cana-de-açúcar e os incêndios de pastagens há o aumento significativo na concentração de materiais particulados na atmosfera (Allen et al., 2004). A elevação dos níveis de materiais particulados na atmosfera de Botucatu é fortalecida pela baixa frequência de precipitação no período decorrente. Destaca-se ainda a presença de duas represas hidroelétricas, Jurumirim (23,15°S; 48,04°W) com reservatório de aproximadamente 449 km² de extensão e Barra Bonita (22,37°S; 48,19°W) que possui 308 km² de extensão. Suas construções representam alterações na cobertura do solo e no microclima local.

6.3.2 Modelos Estatísticos

O primeiro modelo estatístico utilizado neste trabalho, afim de comparação, foi baseado na relação entre a fração transmitida da irradiação direta na incidência normal (kt_b) [razão entre H_b e a irradiação direta integrada no topo da atmosfera (H_{sc}^d)] com a transmissividade atmosférica da irradiação global (kt) [razão entre a H_g e a irradiação global no topo da atmosfera (H_0)] nas partições horárias e diárias (LIU:JORDAN, 1960). No segundo modelo, kt_b é relacionado com a razão de insolação ($r' = n/N$), onde n representa a insolação e N o fotoperíodo local diário em

horas (ÅNGSTRÖM, 1924). Neste trabalho os modelos estatísticos ajustados com dados locais são comparados com modelos de Rede Neural Artificial. Na partição horária a fração transmitida é obtida pela razão entre H_b horária e a constante solar integrada na hora ($4,921 \text{ MJm}^{-2}$).

Tabela 2. Modelos estatísticos utilizados na comparação.

Relação de kt_b vs kt		
Partição	Símbolo	Modelo
Horário	ME1	$kt_b = -0,035 + 0,775 \times kt - 4,805 \times kt^2 + 12,503 \times kt^3 - 7,906 \times kt^4$
Diária	ME2	$kt_b = -0,123 + 1,768 \times kt - 8,213 \times kt^2 + 16,617 \times kt^3 - 9,584 \times kt^4$
Relação de kt_b vs n/N		
Diário	ME3	$kt_b = 0,007 + 0,239 \times (n/N) + 0,285 \times (n/N)^2$

6.3.3 Redes Neurais Artificiais (RNAs)

As RNAs foram usadas com sucesso na resolução de problemas complexos em diferentes áreas de aplicação, incluindo o reconhecimento de padrões, identificação, classificação e regressão (HASNI et al., 2012). As RNAs são sistemas computacionais de minerações de dados maciçamente paralelo distribuído, constituído de unidades de processamentos simples que consistem num grupo interconectado de neurônios artificiais (HAYKIN, 1998). As RNAs são capazes de realizar análise preditiva de dados, armazenando o conhecimento e solucionando problemas de complexa relação não-linear entre os dados de saída e os de entrada, dos quais os modelos estatísticos são limitados (BEHRANG et al., 2010; MARTÍ:GASQUE, 2011). As RNAs são compostas de unidades de processamento simples e, normalmente, são estruturadas em três camadas distintas: entrada que apresenta os dados selecionados para o modelo, uma ou mais camadas ocultas que processam os dados e a camada de saída que produz os resultados (LORENA et al., 2011). Dentre as diferentes tipos de RNAs, destaca-se a Perceptron de Múltiplas Camadas (MLP).

A Perceptron de Múltiplas Camadas (MLP), que é amplamente utilizado em modelagem para resolução de problemas complexos, consiste de um conjunto de unidades sensoriais (nós de fonte) com três camadas interligadas. A primeira camada é a entrada ($x_{i,j}$), a segunda possui uma ou mais camadas ocultas de nós

computacionais com pesos de ligações ($w_{i,j}$) e a terceira camada corresponde à saída de nós computacionais (y_i) (Figura 2).

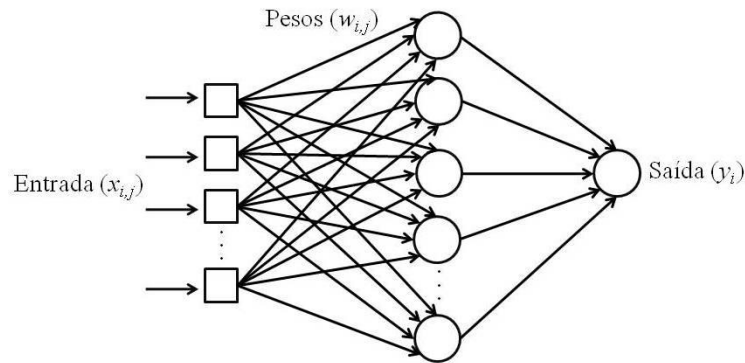


Figura 2. Arquitetura de uma MLP com uma camada oculta.

A saída da rede (y_i) pode ser modelada conforme (LAM et al., 2008):

$$y_i = \sum_{j=1}^n w_{i,j} x_{i,j} + \theta_i \quad (1)$$

onde $x_{i,j}$ é o sinal de entrada, a partir de neurônios de ordem j (para a camada de entrada), $w_{i,j}$ é o peso da ligação direta do neurônio j para neurônio i (na camada oculta) e θ_i é a bias (erro) do neurônio i .

Na MLP, o sinal de entrada se propaga para frente através da rede, camada por camada. Dentro de cada camada há vários neurônios (nós), que estão ligados aos nodos na camada adjacente com pesos diferentes. Cada entrada é multiplicada por um peso de ligação. Após o treinamento os pesos terão informações importantes, antes eles são aleatórios e não têm nenhum significado. A saída dos neurônios é calculada pela aplicação de uma função de ativação. A função de ativação utilizada é tipicamente sigmóide padrão (Eq.2).

$$f(x) = \frac{1}{[1+\exp(-x)]} \quad (2)$$

Existem muitos algoritmos de aprendizado, específicos para determinados modelos de redes neurais. Neste trabalho a MLP foi treinada usando o algoritmo de aprendizado *backpropagation*. O aprendizado consiste em duas passagens

pelas diferentes camadas das redes: um passe para frente, a propagação, e um passe para trás, a retropropagação (HAYKIN, 1998). No passo para frente, um padrão de atividade (entrada de sinal vetorial) é aplicado aos nós sensoriais da rede e seus efeitos se propagam através da rede de camada por camada (HAYKIN, 2001). O termo momento é utilizado para aumentar a taxa de aprendizado. Neste trabalho os modelos gerados são simbolizados com RNA seguido de uma numeração. O modelo horário com kt é simbolizado com (RNA1), o modelo diário com kt como variável de entrada (RNA2) e o modelo diário com a razão de insolação como variável de entrada (RNA3).

6.3.4 Software e metodologia utilizada

A Waikato Environment for Knowledge Analysis (WEKA) é um conjunto de algoritmos de aprendizado de máquinas para tarefas de mineração de dados, disponível em: <http://www.cs.waikato.ac.nz/~ml/weka/> (Hall et al., 2009). Fornecendo ao usuário uma programação em Java, o WEKA contém ferramentas para pré-processamento de dados, classificação, regressão, regras de associação e visualização (OLIVEIRA et al., 2006; LORENA et al., 2011). A MLP foi treinada através da ferramenta computacional WEKA. Várias simulações foram realizadas com o número de camadas ocultas e diferentes números de neurônios, mas a porcentagem de erro encontrado não mostrou alteração significativa. Assim, o número de camadas utilizadas foi o padrão disponível no WEKA, definido como “a” = (número de variável de entrada + número de Classes)/2, assim como o número de neurônios da rede. Os parâmetros Learning rate, o Momentum e o Training Time utilizados foram, respectivamente, 0,3; 0,2 e 500. Maiores detalhes sobre mineração de dados e o WEKA são encontrados em (WITTEN et al., 2011).

As correlações estabelecidas para os modelos de RNAs têm as mesmas variáveis de entrada dos modelos estatísticos. Neste trabalho a entrada (x_{ij}) para as RNAs foram o kt e a razão de insolação (n/N), a saída (y_i) é a fração transmitida (kt_b), que em seguida é convertida na irradiação direta na incidência normal (H_b).

6.3.5 Cálculo da H_0 , do fotoperíodo e H_{sc}^d

A irradiação global no topo da atmosfera (H_0) foi calculada de acordo com equações descritas em (SOUZA et al., 2005), conforme Equação (3):

$$H_0 = I_0 \left[1 + 0,033 \cos \left(\frac{2\pi DJ}{365} \right) \right] \times [(\sin \varphi \sin \delta) + (\cos \varphi \cos \delta \sin \omega_s)] \quad (3)$$

em que I_0 é a constante solar ($= 1367 \text{ Wm}^{-2}$), DJ é o dia juliano começando em 1 de janeiro (para primeiro de janeiro, $DJ = 1$, e para 31 de dezembro, $DJ = 365$), φ é a latitude local (em graus), δ é a declinação solar (em graus) Eq. (4), ω_s é o ângulo horário (em graus) Eq.(5).

$$\delta = 23,45 \times \sin \left[\frac{360(DJ+284)}{365} \right] \quad (4)$$

$$\omega_s = \cos^{-1}[-\tan(\delta) \times \tan(\varphi)] \quad (5)$$

O fotoperíodo (N) é calculado conforme a Eq.(6):

$$N = \frac{2}{15} \times \omega_s \quad (6)$$

A irradiação direta no topo da atmosfera é obtida multiplicando a constante solar integrada ($4,921 \text{ MJm}^{-2}$) pelo fotoperíodo (Eq. 7):

$$H_{sc}^d = 4,921 \times N \quad (7)$$

6.3.6 Base de dados para validação

A série climática foi utilizada para treinamento e validação da rede neural. Foi criado um script para a separação dos dados. O script faz a leitura dos cinco primeiros valores, separa os quatro primeiros para o treinamento e o último para validação. Em seguida faz a leitura dos cinco valores sequenciais e executa o mesmo procedimento anterior. Este processamento é realizado automaticamente até a separação total dos dados para treinamento e validação. Como resultado final, 80% (37.774 horas e 3.444 dias) da série foi utilizada para treinamento e 20% (9.443 horas e 860 dias) para validação. Para efeito de comparação, os modelos estatísticos foram validados com a mesma base de validação das RNAs.

6.3.7 Avaliação estatística

Na avaliação do desempenho estatístico dos modelos foram utilizados: *Mean Bias Error* (MBE), *Relative Mean Bias Error* (rMBE), *Root Mean Square Error* (RMSE), *Relative Root Mean Square Error* (rRMSE) e *d* de Willmott (SANTOS et al., 2014b):

$$\text{MBE (MJm}^{-2}) = \frac{\sum_{i=1}^{N'} (P_i - O_i)}{N'} \quad (6)$$

$$\text{rMBE (\%)} = 100 \times \frac{\frac{\sum_{i=1}^{N'} (P_i - O_i)}{N'}}{\bar{X}} \quad (7)$$

$$\text{RMSE (MJm}^{-2}) = \left[\frac{\sum_{i=1}^{N'} (P_i - O_i)^2}{N'} \right]^{\frac{1}{2}} \quad (8)$$

$$\text{rRMSE (\%)} = 100 \times \frac{\left[\frac{\sum_{i=1}^{N'} (P_i - O_i)^2}{N'} \right]^{\frac{1}{2}}}{\bar{X}} \quad (9)$$

$$d = 1 - \frac{\sum_{i=1}^{N'} (P_i - O_i)^2}{\sum_{i=1}^N (|P'_i| + |O'_i|)^2} \quad (10)$$

em que: P_i representa os valores estimados de H_b , O_i os valores medidos de H_b , $|P'_i|$ o valor absoluto da diferença $P_i - \bar{O}_i$, em que $\bar{O}_i$, representa a média de O_i , $|O'_i|$ representa o valor absoluto da diferença $O_i - \bar{O}_i$. $\bar{X}$ é o valor médio medido e N' o número de observações. Valores positivos (negativos) de MBE (rMBE) indicam superestimativa (subestimativa). O RMSE (rRMSE) dá precisão a curto prazo das estimativas com relação aos dados medidos. Diferentes intervalos de rRMSE são definidos para avaliar a acurácia dos modelos (Jamieson et al., 1991; Heinemann et al., 2012):

Excelente se (excellent if) $\text{rRMSE} < 10\%$;

Bom se (good if) $10\% \leq \text{rRMSE} < 20\%$;

Aceitável se (fair if) $20\% \leq \text{rRMSE} < 30\%$;

Pobre se (poor if) $\text{rRMSE} \geq 30\%$.

O índice de ajustamento “d” varia de 0 a 1 e representa a total discordância ou a concordância perfeita, respectivamente, entre valores medidos e

estimados (WILLMOTT, 1982). A aplicação conjunta de MBE, RMSE e “d” na avaliação dos modelos possibilita uma análise ampla do desempenho dos modelos.

6.4 RESULTADOS E DISCUSSÃO

Inicialmente foi estimado kt_b com RNAs. Para obtenção de H_b os valores kt_b estimados são multiplicados pela irradiação direta no topo da atmosfera. Os resultados obtidos com as RNAs e com os modelos estatísticos são apresentados em quatro subseções. Na primeira, as estimativas dos modelos horário e diário em função de kt são analisadas. Na segunda subseção são analisados os modelos diários da razão de insolação como variável de entrada. A comparação entre o desempenho da RNA com relação aos modelos estatísticos é realizada na terceira subseção. Na quarta subseção um estudo de caso é realizado para verificar o efeito dos aerossóis e do vapor d’água na estimativa.

6.4.1 Estimativas com modelos baseados na transmissividade atmosférica (kt)

Os resultados mostram que os modelos RNA1 e RNA2 têm bom desempenho estatístico (Tabela 3). O modelo RNA1 teve: $MBE = -0,16 \text{ MJm}^{-2}$, $RMSE = 0,45 \text{ MJm}^{-2}$ ($rRMSE = 22,20\%$) e $d = 0,94$. Os erros gerados pelo modelo ME1 são melhores que os obtidos com o RNA1. Os modelos ME1 e ME2 tem valores de $rRMSE$ iguais a 21,98 e 20,13%, respectivamente. Os modelos da partição horária subestimam as medidas. A obtenção de H_b a partir de kt_b é de interesse por melhorar os resultados na correlação entre H_b e H_g , que normalmente é baixa.

Tabela 3. Índices estatísticos de validação dos modelos.

Tempo	Modelo	MBE (MJm^{-2})	rMBE (%)	RMSE (MJm^{-2})	rRMSE (%)	d
Horário	RNA1	-0,16	-8,08	0,45	22,20	0,94
	ME1	-0,04	-2,24	0,44	21,98	0,94
Diário	RNA2	-0,47	-2,90	2,70	16,52	0,98
	ME2	0,09	0,57	3,27	20,13	0,97

Os modelos horários tiveram valores médios de rRMSE e d iguais a 22,09% e 0,94, respectivamente. Estes resultados comprovam que o modelo gerado pode ser utilizado para estimar H_b^h com bons resultados. Os modelos estatísticos, também estimaram H_b^h com erros aceitáveis. Por ser uma técnica AM, aqui a RNA1 é indicada como a primeira escolha. Os resultados obtidos neste trabalho são melhores que os encontrados por, Polo et al., (2011), que propuseram um modelo para geração de dados de H_b e obtiveram rRMSE de 31%. As usinas solares necessitam de dados de H_b em intervalo de tempo menor, assim, a estimativa horária fornece informações importantes para projetos de usinas solarimétricas.

Na estimativa diária o resultado com RNA2 é melhor que do modelo estatístico: MBE = -0,474 MJm⁻², RMSE = 2,701 MJm⁻² (rRMSE = 16,519%) e d = 0,980. Similarmente ao modelo horário, na modelagem diária o RNA2 é indicado como a primeira escolha. Os modelos de redes neurais estimam com precisão igual ou superior aos modelos estatísticos. Esta observação comprova que os modelos gerados aqui são boas metodologias na estimativa de H_b^d .

A Figura (3a) apresenta a comparação entre os valores medidos e estimados de H_b^h para RNA1, enquanto na Figura (3b), são mostrados os resultados do modelo ME1. Nota-se que, RNA1 apresenta desempenho similar aos modelos estatísticos. Os pontos acompanham a linha central de comparação. Embora o valor esperado de H_b para uma hora totalmente de céu limpo seja de $\approx 4,92$ MJm⁻² (constante solar integrada), a rede neural não estimou valores maiores que 3,35 MJm⁻². Este resultado pode ser causado pela atenuação da radiação (por aerossóis e absorção pelo vapor d'água). Assim como, pela variável de entrada kt não apresentar variações temporais similares a kt_b . Normalmente, nos modelos da partição horária é esperada maior dispersão dos pontos que nas estimativas de partição maior.

Na Figura 3a, a maioria dos pontos está abaixo da linha ideal. No ME1 62,34% das estimativas estão compreendidas no intervalo de 1 a 3 MJ m⁻². Semelhante ao RNA1, o modelo ME1 tem a maioria dos pontos abaixo da linha ideal de comparação. Os modelos podem ser utilizados por engenheiros, projetistas e pesquisadores de recursos solares na estimativa de H_b^h . Os valores de R^2 entre as regressões não diferem estatisticamente. O coeficiente de determinação (R^2) gerado, em média, explica 79,5% da relação entre H_b^d medida com a estimada.

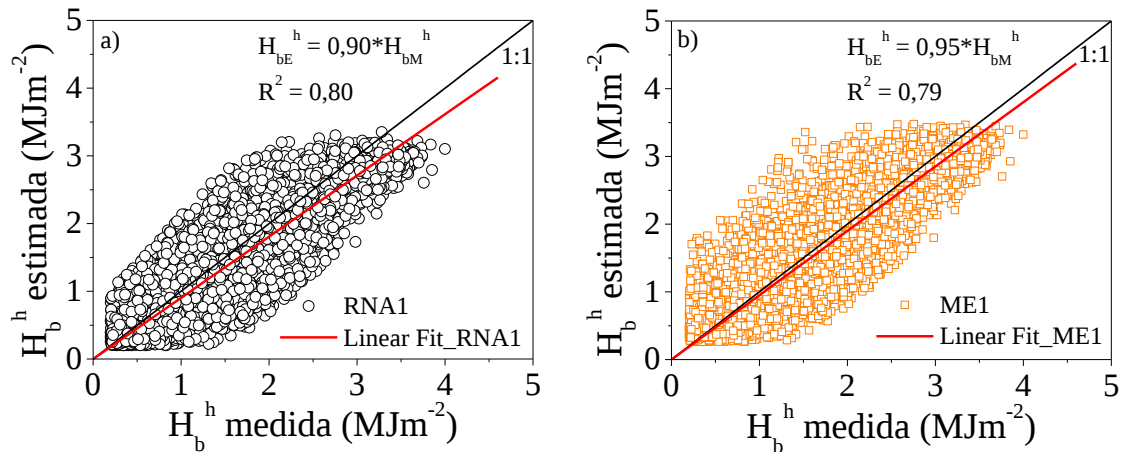


Figura 3. Dispersão entre os valores medidos e estimados dos modelos horários a partir da relação entre $kt_b \times kt$, a) RNA1 e b) ME1. H_{bE}^h é a irradiação direta na incidência normal estimada; H_{bM}^h é a irradiação direta na incidência normal medida.

Os diagramas de dispersão do modelo de rede RNA2 e do modelo ME2 para estimativa de H_b^d são mostrados nas Figuras 4 (a, b), respectivamente. Observa-se que, a dispersão é similar entre RNA2 e ME4. Mas, a estimativa com RNA2 tem melhor ajuste sobre linha de comparação. Com base nas curvas de dispersão pode-se destacar que os modelos RNA2 e ME2 podem ser utilizados para estimar H_b^d com boa precisão. Contudo, a técnica de AM tem melhor capacidade de adequação.

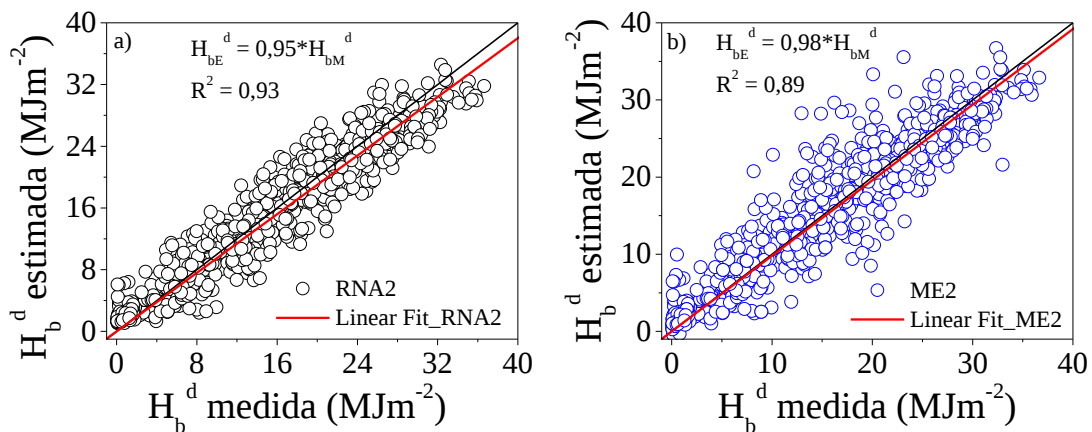


Figura 4. Dispersão entre os valores medidos e estimados dos modelos diário de H_b^d obtidos a partir da relação entre $kt_b \times kt$. a) RNA2 e b) ME2. H_{bE}^d é a irradiação direta na incidência normal estimada; H_{bM}^d é a irradiação direta na incidência normal medida.

É observado maior valor de R^2 é para o modelo RNA2, comprovando o potencial da rede neural em estimar H_b^d . Ressalta-se que nem sempre elevado coeficiente de correlação indica boa precisão do modelo em estimar determinada variável de interesse (Willmontt, 1982). Analisar o desempenho da rede neural nas diferentes condições de coberturas de nuvens é interessante na verificação da relação entre as condições de céu predominantes e os erros gerados na estimativa. Nos resultados locais, as nuvens são responsáveis pela atenuação de H_b , em sua ausência os aerossóis e o vapor d'água são os principais fatores atenuantes.

6.4.2 Estimativas dos modelos baseados na razão de insolação

Os resultados estatísticos (rMBE, rRMSE e d) de validação dos modelos estão presentes nas Figs. 5 (a, b, c). Os desempenhos dos modelos que correlacionam kt_b com a razão de insolação são piores que dos modelos que correlacionam kt_b com kt . O mais provável é que estas diferenças sejam causadas pelos erros intrínsecos nas medidas de n (erro de leitura do observador e erro na queima da heliograma, por exemplo). Os RNA3 e ME3 tiveram boa precisão na estimativa, com resultados estatísticos similares entre si. O RNA3 tem rMBE = -6,947%, rRMSE = 20,068% e d = 0,941. O ME3 estima com rMBE = -4,456%, rRMSE = 19,273% e d = 0,947. O índice “d” para ME3 foi de 0,918. Por isso, redes neurais e o ME3 são recomendadas para estimativa de H_b^d nos locais que não dispõe de pireliômetro e dispõe de dados de insolação solar. Aqui, o RNA3 é indicado como a primeira escolha.

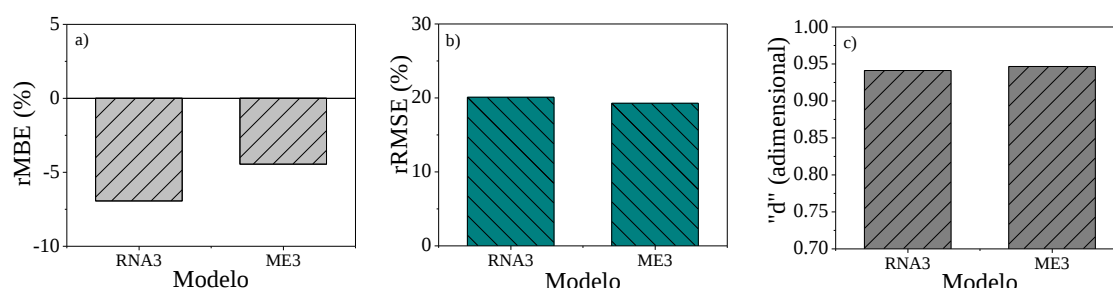


Figura 5. Índices estatísticos de validação dos modelos diários de H_b^d obtidos a partir da relação entre kt_b x n/N . a) rMBE, b) rRMSE e c) “d” de Willmontt.

A H_b^d estimada é representada nas Figs. 6 (a, b) na forma de gráficos de dispersão para os modelos. A partir destes confirma-se a boa estimativa de

H_b^d . Acima de $26,00 \text{ MJm}^{-2}$ os valores estimados subestimam as medidas. Estes resultados confirmam as observações analisadas com os índices estatísticos. O coeficiente de determinação médio indica que mais de 82% da variação de H_{bE}^d é explicada por H_{bM}^d . Na prática, os modelos desenvolvidos aqui podem ser utilizados para estimativa de H_b com resultados satisfatórios.

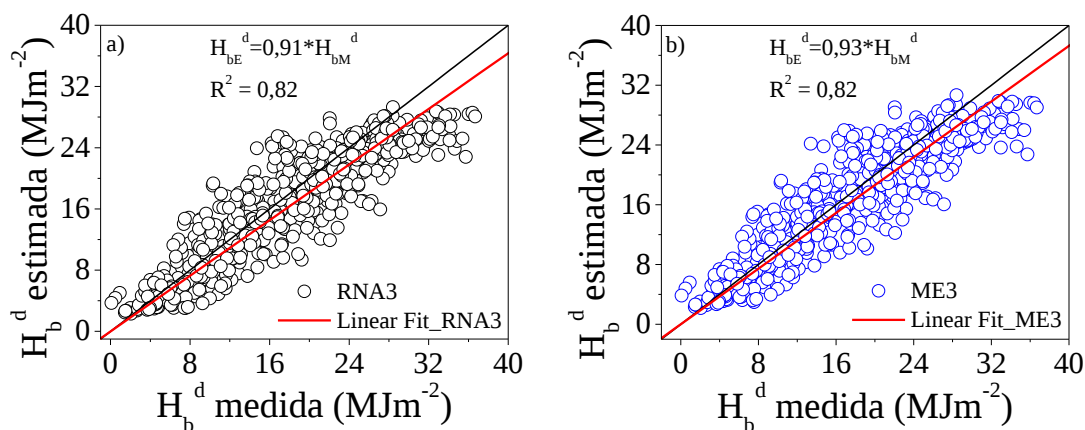


Figura 6. Dispersão entre os valores medidos e estimados dos modelos diários a partir da relação entre $kt_b \times n/N$, a) RNA3 e b) ME3. H_{bE}^d é a irradiação direta na incidência normal estimada; H_{bM}^d é a irradiação solar direta na incidência normal medida.

6.4.3 Comparação das redes neurais com os modelos estatísticos

Originalmente o presente trabalho avaliou a RNA na estimativa de H_b . A variável de entrada também foi avaliada, para mostrar a correlação mais relevante na estimativa H_b . O rRMSE médio obtido com os três modelos de rede neural (RNA1, RNA2 e RNA3) é de $19,596 \pm 2,870\%$ (desempenho bom), com os modelos estatísticos é de $20,461 \pm 1,384\%$ (desempenho aceitável). Os valores de rRMSE das redes neurais são iguais ou inferiores aos dos modelos estatísticos, que confirmam a capacidade da RNA na modelagem. O d médio obtido com RNA é de $0,952 \pm 0,020$, com modelos estatísticos é de $0,952 \pm 0,016$. Os valores de R^2 dos modelos de RNA são semelhantes ou mais altos quando comparados aos modelos estatísticos. Em geral, o desempenho dos modelos de RNA é superior aos modelos estatísticos. Estes resultados expressam a boa generalização da rede neural, e permite a utilização nas estimativas confiáveis de H_b , gerando mapas de irradiação direta na incidência normal para aplicação nas diferentes áreas.

6.4.4 Estimativa de H_b com modelos de redes neurais: Estudo de caso para condição de céu sem nuvens

A estimativa na condição de céu sem nuvens é realizada para verificar o efeito dos aerossóis e do vapor d'água na modelagem de H_b . Os dados utilizados foram para $kt > 0,65$. Foi utilizada a divisão percentual (*Percentage Split*) do WEKA, com 70% dos dados utilizados para treinamento e 30% para validação. A escolha dos dados para treinar e validar é feita aleatoriamente. A porcentagem de dados para treinamento e validação é diferente da primeira parte dos resultados, por se tratar de dados apenas de céu sem nuvens e quantidade de dados reduz significativamente com relação à base total. Na condição de céu sem nuvens RNA1, RNA2, RNA3 superestimaram e tiveram espalhamento menor que na validação com todas as condições de cobertura de nuvens (Figura 7a, b, c). O RNA1 teve $rRMSE = 12,723\%$, RNA2 com $rRMSE = 12,813\%$ e RNA3 resultou em $rRMSE = 11,308\%$. Os RNA1 e RNA2 tiveram as maiores superestimativas. O modelo RNA3 gerou o menor $rMBE$ (0,526%). Isto é um indicativo que na condição de céu claro as medidas da insolação estão sujeitas a menores erros na queima do heliograma. Os valores de d foram baixos. Climaticamente, o resultado mostra que as nuvens são responsáveis por aumentarem consideravelmente o espalhamento.

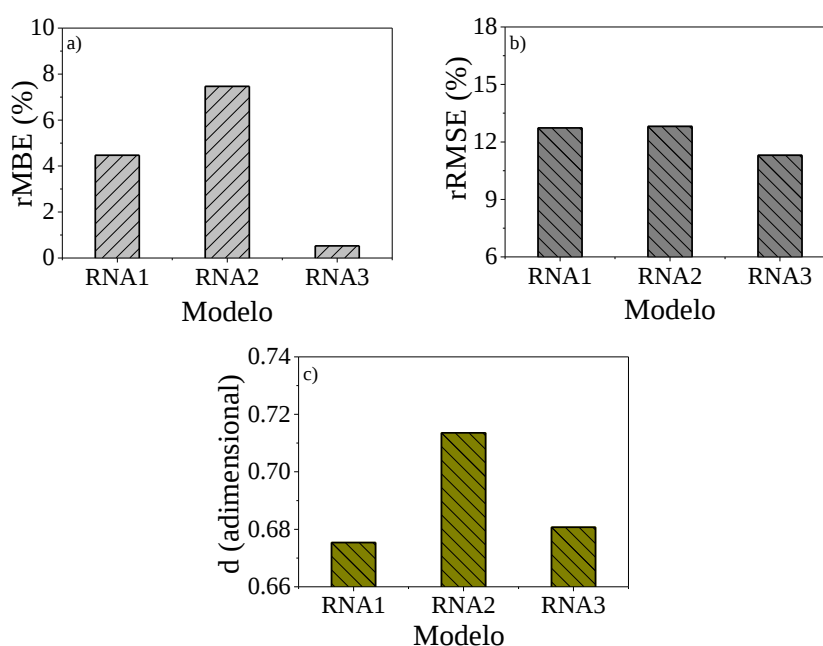


Figura 7. Índices estatísticos de validação de RNA1, RNA2 e RNA3 para condição de céu claro. a) $rMBE$, b) $rRMSE$ e c) “ d ” de Willmontt.

Na ausência de nuvens, as variabilidades dos valores de rRMSE e d estão relacionadas com a flexibilidade da rede neural em estimar H_b e muito provavelmente ao efeito do espalhamento causado pelos aerossóis e absorção do vapor d'água. A explicação provável resume ao fato de existir um grande número de combinação de aerossóis e vapor d'água que irão gerar valores de kt no mesmo intervalo de variação, mas não se pode esperar que a mesma combinação produza um valor único de kt_b ao ser modelado. Este comportamento foi observado para a região de Botucatu, que é influenciada pelas queimadas agrícolas (cana de açúcar e florestas) e por porções de água que se situam próximas ao local de estudo. Estas duas são responsáveis pela emissão de aerossóis e vapor d'água, respectivamente.

A dispersão para condição de céu sem nuvens é caracterizada por ser menor que na condição de todos os tipos de cobertura de céu (Figura 8a, b, c). Quando comparados RNA1 e RNA2 com RNA3, nota-se que o efeito dos aerossóis e do vapor d'água são mais significativos nos modelos baseados em kt . Este resultado é quantificado pelos valores de rMBE e rRMSE e qualificado pela curva de dispersão. Quando são observados valores baixos de H_b associados com estimativas mais elevadas, a mesma explicação para a variabilidade de rRMSE também é válida aqui.

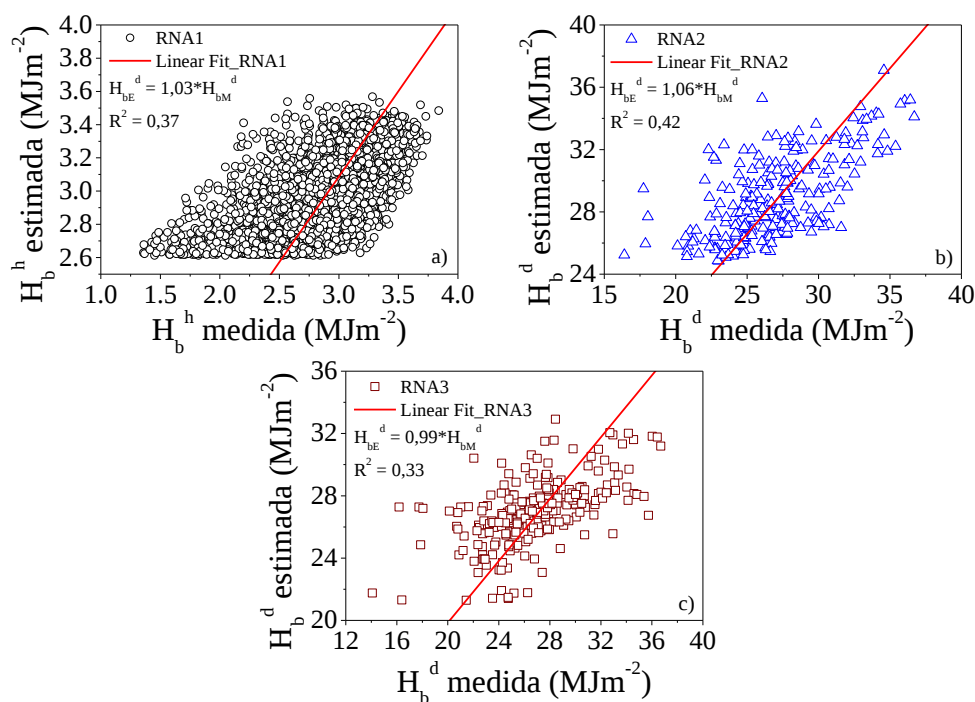


Figura 8. Dispersão entre os valores medidos e estimados para condição de céu claro. a) RNA1, b) RNA2 e c) RNA3.

6.5 CONCLUSÕES

Os desempenhos de modelos gerados pela RNA e dos modelos estatísticos na estimativa da irradiação direta na incidência normal, usando a k_t e a n/N como variáveis de entrada, são avaliados na região de Botucatu – SP. Foram utilizados três modelos estatísticos ajustados com dados locais. A rede neural utilizada foi a MLP.

As variáveis de entrada utilizadas são capazes de explicar a maior parte da variabilidade da irradiação direta na incidência normal em Botucatu. Os modelos na partição horária tiveram desempenho piores quando comparados com os modelos diários. As grandes dispersões observadas nos modelos foram devidas a nebulosidade. A nebulosidade foi responsável pelo aumento no rRMSE em 9,473% no RNA1; 3,706% com RNA2 e 8,760% com RNA3. No estudo de caso para céu limpo, a dispersão foi causada, principalmente, pela atenuação da radiação pelos aerossóis e pela absorção por vapor d'água.

O modelo da rede neural utilizado teve desempenho superior aos modelos estatísticos e deve ser a primeira escolha na estimativa de H_b . A estimativa de H_b foi satisfatória e potencialmente para ser empregada com sucesso nos locais com escassez de informações solarimétrica.

A comparação da rede neural com modelos físicos de transferência radiativa deve ser considerada em trabalhos futuros. A utilização de novas variáveis de entrada, ainda não avaliadas, pode ser investigada na modelagem de H_b . A rede pode ser utilizada para avaliar seu potencial na modelagem de outras radiações. Outros modelos ou técnicas de aprendizado de máquinas podem ser desenvolvidos e utilizados para modelagem de H_b . Projetos de usinas solares necessitam de informações solarimétricas em menores partições de tempo, neste caso, o desenvolvimento de novos modelos de redes neurais para estimava de H_b em partições de tempo menores (10 e 5 min, por exemplo) devem ser considerados em estudos futuros.

6.6 REFERÊNCIAS BIBLIOGRÁFICAS

ALLEN, A. G.; CARDOSO, A. A.; DA ROCHA, G. O. Influence of sugar cane burning on aerosol soluble ion composition in Southeastern Brazil. **Atmospheric Environment**, v. 38, p. 5025 – 5038, 2004.

ÅNGSTRÖM, A. K. Solar and terrestrial radiation. **Quarterly Journal of the Royal Meteorological Society**, v.50, p.121, 1924.

BEHRANG, M. A.; ASSAREH, E.; GHANBARZADEH, A.; NOGHREHABADI, A. R. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. **Solar Energy**, v.84, p.1468-1480, 2010.

CODATO, G.; OLIVEIRA, A. P.; SOARES, J.; ESCOBEDO, J. F.; GOMES, E. G.; DAL PAI, A. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. **Theoretical and Applied Climatology**, v. 93, p.57–73, 2008.

ESCOBEDO, J. F.; GOMES, E. N.; OLIVEIRA, A. P.; SOARES, J. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. **Renewable Energy**, v.36, p.169-178, 2011.

GUEYMARD, C. A. Direct solar transmittance and irradiance predictions with broadband models. Part I: detailed theoretical performance assessment. **Solar Energy**, v.74, 355–379, 2003a.

GUEYMARD, C. A. Direct solar transmittance and irradiance predictions with broadband models. Part II: validation with high-quality measurements. **Solar Energy**, v.74, 381–395, 2003b.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. **The WEKA Data Mining Software: An Update**; SIGKDD Explorations, v.11, Issue 1, 2009.

HASNI, A.; SEHLI, A.; DRAOUI, B.; BASSOU, A.; AMIEUR, B. Estimating global solar radiation using artificial neural network and climate data in the south-western region of Algeria. **Energy Procedia**, v.18, p.531 – 537, 2012.

HAYKIN, S. **Neural networks: A comprehensive foundation**. 2nd ed. Hamilton: Prentice Hall, 1998, 897p.

HAYKIN, S. **Redes Neurais: Princípios e prática**. Trad. Paulo Martins Engel. – 2. Ed. – Porto Alegre: Bookman, 2001, 900p. [in portuguese].

HEINEMANN, A. B.; VAN OORT, P. A. J.; FERNANDES, D. S.; MAIA, A. H. N. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. **Bragantia, Campinas**, v. 71, n. 4, p.572-582, 2012.

JAMIESON, P. D.; PORTER, J. R.; WILSON, D. R. A test of the computer simulation model ARC-WHEAT1 on wheat crops grown in New Zealand. *Field Crops Research*, v.27, p.337-350, 1991.

KALOGIROU, S. A.; MATHIOULAKIS, E.; BELESSIOTIS, V. Artificial neural networks for the performance prediction of large solar systems. *Renewable Energy*, v.63, p.90-97, 2014.

KOCA, A.; OZTOP, H. F.; VAROL, Y.; KOCA, G. O. Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. *Expert Systems with Applications*, v.38, p.8756-8762, 2011.

LAM, J. C.; WAN, K. K. W.; LIU, Y. Solar radiation modeling using ANNs for different climates in China. *Energy Conversion and Management*, v.49, p.1080–1090, 2008.

LIU, B. Y. H.; JORDAN, R. C. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. **Solar Energy**, v.3, p.1-19, 1960.

LOPEZ, G.; RUBIO, M. A.; BATLLES, F. J. Estimation of hourly direct normal from measured global solar irradiance in Spain. *Renewable Energy*, v.21, p. 175-186, 2000.

LORENA, A. C.; JACINTHO, L. F. O.; SIQUEIRA, M. F.; GIOVANN, R.; LOHMANN, L. G.; CARVALHO, A. C.P.L.F.; YAMAMOTO, M. Comparing machine learning classifiers in potential distribution modelling. **Expert Systems with Applications**, v.38, p.5268–5275, 2011.

LYRA, G. B.; ZANETTI, S. S.; SANTOS, A. A. R.; SOUZA, J. L.; LYRA, G. B.; OLIVEIRA-JÚNIOR, J. F.; LEMES, M. A. M. Estimation of monthly global solar irradiation using the Hargreaves–Samani model and an artificial neural network for the state of Alagoas in northeastern Brazil. *Theoretical and Applied Climatology*. DOI 10.1007/s00704-015-1541-8, 2015.

MANZANO, A.; MARTÍN, M. L.; VALERO, F.; ARMENTA, C. A single method to estimate the daily global solar radiation from monthly data. *Atmospheric Research*, v.166, p.70–82, 2015.

MARTÍ, P.; GASQUE, M. Improvement of temperature-based ANN models for solar radiation estimation through exogenous data assistance. *Energy Conversion and Management*, v.52, p.990-1003, 2011.

NFAOUI, H.; BURET, J. Estimation of daily and monthly direct, diffuse and global solar radiation in Rabat (Morocco). *Renewable Energy*, v. 3, p. 923-930, 1993.

OLIVEIRA, A. L. I. Estimation of software project effort with support vector regression. *Neurocomputing*, v.69, p. 1749-1753, 2006.

POLO, J.; ZARZALEJO, L. F.; MARCHANTE, R.; NAVARRO, A. A. A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. *Solar Energy*, v.85, p.1164–1170, 2011.

QAZI, A.; FAYAZ, H.; WADI, A.; RAJ, R. G.; RAHIM, N. A.; KHAN, W. A. The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. *Journal of Cleaner Production*, v.104, p.1-12, 2015.

SANTOS, C. M, SOUZA, J. L, FERREIRA JUNIOR, R. A.; TIBA, C.; MELO, R. O.; LYRA, G. B.; TEODORO, I.; BASTOS LYRA, G.; LEMES, M. A. M. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. *Energy*, v.71, p.388-398, 2014a.

SANTOS, C. M.; SOUZA, J. L.; TERAMOTO, E. T.; TIBA, C.; MELO, R. O. Modelagem da irradiação solar global média horária mensal (H_g^h) para quatro localidades de Alagoas/Brasil. *Nativa*, Sinop, v.2, p. 79-88, 2014b. [in portuguese].

SOARES, J.; OLIVEIRA, A. P.; BOZNAR, M. Z.; MLAKAR, P.; ESCOBEDO, J. F.; MACHADO, A. J. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique. *Applied Energy*, v.79 201–214, 2004.

SOUZA, J. L.; NICÁCIO, R. M.; MOURA, M. A. L. Global solar radiation measurements in Maceió, Brazil. *Renewable Energy*, v. 30, p. 1203 – 1220, 2005.

VIGNOLA, F.; MCDANIELS, D. K. Beam-global correlations in the Pacific Northwest. *Solar Energy*, v. 36, p. 409-418, 1986.

WANA, K. K. W.; TANGA, H. L.; YANG, L.; LAM, J. C. An analysis of thermal and solar zone radiation models using an Angstrom–Prescott equation and artificial neural networks. *Energy*, v.33, p.1115–1127, 2008.

WANGA, J.; WANG, E.; YIN, H.; FENG, L.; ZHAO, Y. Differences between observed and calculated solar radiations and their impact on simulated crop yields. *Field Crops Research*, v.176, p.1–10, 2015.

WILLMOTT, C. J. On the validation of models. *Physical Geography, Delaware*, v. 2, p. 184-194, 1982.

WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd ed. 2011, 630p.

WORLD METEOROLOGICAL ORGANIZATION (WMO). *Guide to the Meteorological Instruments and Methods of Observation*. WMO-8; 1983. p. 517.

YADAV, A. K.; CHANDEL, S. S. Solar radiation prediction using Artificial Neural Network techniques: A review. *Renewable and Sustainable Energy Reviews*, v.33, p.772–78, 2014.

7. VARIABILIDADE TEMPORAL DA TURBIDEZ ATMOSFÉRICA E ATENUAÇÃO DE DNI EM REGIÃO DE CANA DE AÇÚCAR - BOTUCATU/SP.

7.1 RESUMO

Neste trabalho a atenuação da irradiância direta na incidência normal (DNI) na região de Botucatu/São Paulo, área sob influência de queimadas agrícolas locais e adjacentes, é expressa utilizando o fator de Turbidez de Linke (TL), no período de 1996-2008. Foram utilizadas duas metodologias, representadas como TL_{Dj} e TL_{Li} . A variabilidade temporal (média horária, por estação do ano e média mensal) é apresentada. A turbidez foi correlacionada com a velocidade do vento e com a temperatura do ar. A distribuição de frequência e a frequência acumulada são analisadas para verificar os níveis de predominância da turbidez na atmosfera local. Informações de profundidade óptica de aerossóis a 550 nm (AOD_{550nm}) e vapor d'água foram obtidos pelo satélite Terra por meio do sensor MODIS. O maior grau de transmissão de DNI é observado na parte da manhã e próximo ao meio dia solar a transmissão é menor (maior o valor da TL). A variabilidade diurna de TL é mais evidente no período mais quente, que no período frio. Maio e julho foram os meses de menor atenuação de DNI (maior transparência atmosférica). A maior atenuação de DNI ocorre na primavera ($TL_{Dj} = 4.22 \pm 0.05$ e $TL_{Li} = 4.65 \pm 0.06$) e no verão ($TL_{Dj} = 4.27 \pm 0.14$ e $TL_{Li} = 4.69 \pm 0.15$). A velocidade do vento e a temperatura do ar correlacionam positivamente com TL. Mais de 28 % das horas de

céu limpo, a turbidez obtida excedeu o valor de 4,0. A região de Botucatu é influenciada pelo vapor d'água e pelos aerossóis de diferentes origens. Este estudo conclui que estes fatores reduzem significativamente a DNI incidente na superfície, com maior transparência atmosférica no período frio e a menor no período quente, que impacta na circulação atmosférica.

Palavras-chave: Transparência atmosférica; Aerossóis; Vapor d'água; poluição; irradiância direta na incidência normal.

7.2 INTRODUÇÃO

A irradiância direta na incidência normal (DNI) possui aplicações indispensáveis na tecnologia de concentradores solares, como forçante climática e controladora da biodiversidade terrestre (BATLLES et al., 2000). Ao atravessar uma atmosfera sem nuvens a DNI é atenuada por dois processos principais: espalhamento pelos aerossóis e a absorção pelo vapor d'água, principalmente nos comprimentos de onda do visível e infravermelho. Os aerossóis são pequenas partículas sólidas ou líquidas em suspensão no ar que segue o movimento de massas de ar e pode ter origem terrestre (queimadas agrícolas, fumaça industrial, tempestades de areia, erupções vulcânicas) e origem marinha, como spray marinho (LOPEZ:BATLLES, 2004). Na condição de céu sem nuvens a concentração de aerossóis e o teor de vapor d'água na atmosfera condicionam a magnitude e a variabilidade de DNI (GUEYMARD, 2012). Os aerossóis e o vapor d'água condicionam o nível de transparência atmosférica e a atenuação causada em DNI é chamada de turbidez.

Na determinação da turbidez atmosférica é necessário medidas detalhados de DNI tanto de banda larga e espectrais (SALAZAR, 2011; ELLOUZ et al., 2013). A aquisição e manutenção desses sensores são quase que inviáveis, devido ao elevado custo de importação e manutenção (DOS SANTOS et al., 2014). Por isso, a turbidez atmosférica normalmente é estimada por modelos paramétricos.

Tipicamente, a turbidez atmosférica na condição de céu sem nuvens é representada através da turbidez de Linke (TL) ou do coeficiente de turbidez de Ångström (β). Pela praticidade de uso e capacidade de representar a espessura óptica de uma atmosfera seca e limpa, a TL é frequentemente utilizada como primeira opção

(ELLOUZ et al., 2013). O cálculo da turbidez é importante na modelagem do clima local, em estudos climáticos, estudos de poluição do ar, medida indireta de DNI, medida indireta da concentração de partículas em suspensão no ar e de vapor d'água que absorve, espalha ou reflete a radiação solar (LÓPEZ:BATLLES, 2004; TRABELSI:MASMOUDI, 2011).

A TL tem sido relatada para inúmeros locais (DIABATÉ et al, 2003; ELTBAAKH et al., 2012; HUSSEIN et al., 2000; FORMENTI et al., 2002), mas escassos são os estudos da variabilidade temporal da turbidez de Linke e a transparência atmosférica no Brasil e, particularmente, para regiões do Estado de São Paulo. Em estudos para a atmosfera do Rio de Janeiro – Brasil, Flores et al., 2015, calcularam a turbidez atmosférica com base em Ångström (β) e mostraram uma diminuição nos meses de verão. Os autores destacaram a grande disponibilidade de vapor de água, resultante de influências marítimas e da grande área vegetal.

Responsáveis por emitirem grande quantidade de aerossóis e gases na atmosfera, a queima de florestas nativas para a produção agropecuária e o uso de práticas agrícolas ainda são atividades comuns no Brasil (OLIVEIRA et al., 2011). O presente trabalho trata da determinação da TL sob condição de céu sem nuvens na região agro-industrial de Botucatu do Estado de São Paulo, onde a economia é baseada no setor industrial, nas atividades agrícolas (produção de açúcar e álcool) e comércio. Duas metodologias de turbidez de Linke (TL_{Dj} e TL_{Li}) são utilizadas e comparadas no intuito de analisar seu efeito e influência no cálculo do nível da turbidez. A correlação da turbidez com a velocidade do vento e a temperatura do ar é apresentada. A análise da frequência de ocorrência e frequência acumulada da turbidez é apresentada para caracterizar os índices de turbidez com a condição atmosférica. Este trabalho agrega extrema importância na verificação da elevação dos níveis de poluição na atmosfera de Botucatu, destacando o período de menor transparência atmosférica (maior turbidez), os prováveis fatores responsáveis pela atenuação de DNI e os prováveis locais de origem. Este artigo, também mostra a tendência da variabilidade da turbidez.

7.3 MATERIAL E MÉTODOS

7.3.1 Descrição do local e clima

Os dados utilizados neste estudo foram obtidos da Estação de Radiometria solar instalada na Faculdade de Ciências Agronômicas/UNESP/SP,

localizada no município de Botucatu. Botucatu ($22^{\circ}53'09''\text{S}$; $48^{\circ}26'42''\text{W}$ e altitude média 786m) é um município brasileiro localizado na região centro-oeste do Estado de São Paulo (Figura 1), possui área total de 1.482,642 km² e população estimada em 2015 de aproximada de 139.483 mil habitantes (IBGE, 2014). O município possui elevado gradiente de altitude, entre 400 a 500 m na região mais baixa (depressão periférica) e entre 700 a 900 na região serrana (Planalto Ocidental). Esta diferença provoca variações na temperatura do ar e nos ventos do município. Distante 221 km do oceano Atlântico e 235 km da Capital São Paulo, com bioma cerrado e de mata atlântica, Botucatu apresenta clima temperado quente (mesotérmico), verão quente e úmido com elevada precipitação e inverno seco (ESCOBEDO et al., 2011).

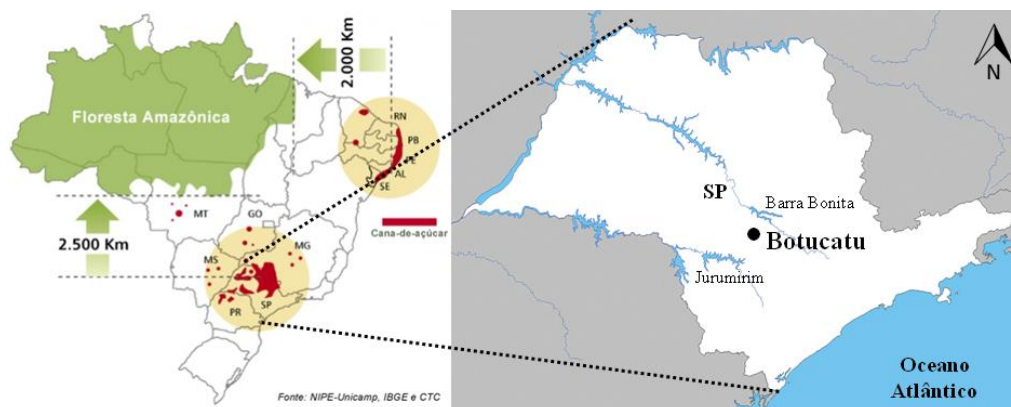


Figura 1. Mapa mostrando as principais áreas de produção de cana no Brasil e municípios no Estado de São Paulo. Adaptado da UNICA <http://www.unica.com.br/production-map/>.

Entre os meses de outubro a março a precipitação é de natureza microclimática originada do processo de convecção livre e de eventos macroclimáticos originados na convergência de massas de ar úmidas oriundas da região amazônica e do oceano atlântico sul, resultando na formação da Zona de Convergência do Atlântico Sul (ZCAS) e nos sistemas frontais (JONES et al., 2004; REBOITA et al., 2010; TERAMOTO:ESCOBEDO, 2012). Entre os meses de junho a setembro, a região central do Brasil é dominada por uma área de alta pressão, pouca precipitação e ventos fracos na baixa troposfera, com convecção na Amazônia deslocada para a parte noroeste da América do Sul (FREITAS et al., 2005; SATYAMURTY et al., 1998). Nos meses da estação seca, principalmente agosto e setembro, é observado a prática da queima de cana-de-açúcar em cidades adjacentes a Botucatu, que são realizadas para a colheita dos

colmos (CODATO et al., 2008). Com a queima de cana-de-açúcar há o aumento significativo na concentração de materiais particulados na atmosfera local (ALLEN et al., 2004). Neste período, destaca-se também uma carga suplementar de aerossóis de queimadas diversas trazido da região Amazônica, da Bolívia, do Paraguai e da Argentina para sudeste do Brasil através da circulação atmosférica (VIDELA et al., 2013; PORTILLO:QUINTERO et al., 2013). Quase todas as queimadas são causadas pelo homem, seja de propósito ou acidental. Assim, o efeito dos aerossóis pode extrapolar a partir de escala local e ser determinante no padrão de redistribuição planetária de energia dos trópicos para médias e altas latitudes através de processos de transporte convectivo (FREITAS et al., 2005), resultando em alterações no ciclo hidrológico, devido as modificações na temperatura atmosférica e na estrutura de nuvens.

Destaca-se ainda a presença de duas represas hidroelétricas, Jurumirim (23,15°S; 48,04°W) e Barra Bonita (22,37°S; 48,19°W), localizadas próxima a Botucatu. A represa de Jurumirim distante em linha reta 70 km de Botucatu, possui reservatório de aproximadamente 449 km² de extensão. A represa de Barra Bonita, distante aproximadamente 30 km de Botucatu, possui 308 km² de extensão. Suas construções representam alterações na cobertura do solo e muito provavelmente no microclima local. Mudanças desse porte ocasionam interferência significativamente nos fluxos turbulentos da atmosfera em micro e mesoescala, modificando o balanço de radiação e a microfísica das nuvens (ROSENFELD, 1999).

7.3.2 BASE DE DADOS E MODELO

7.3.2.1 Descrição das medidas

No cálculo de TL, as medidas de DNI foram obtidas do período de fevereiro de 1996 a dezembro de 2008, através de um pireliômetro da Eppley NIP acoplado a um rastreador solar ST3 da Eppley. A irradiância solar global (R_G) foi monitorada por um piranômetro Eppley PSP e irradiância solar difusa (R_D) por um piranômetro Eppley PSP com um anel de dimensões (40 cm de raio e 10 cm de largura) (DAL PAI et al., 2014). Todos os sensores Foram conectados ao Datalogger CR1000, 23X e CR3000 da Campbell programado para armazenar médias a cada 5 min. Os

sensores são aferidos anualmente pelo método comparativo (WMO, 2008). Foram utilizados valores horários de céu sem nuvens para DNI.

Valores da profundidade óptica dos aerossóis a 550 nm (AOD_{550nm}) e vapor d'água foram utilizados neste trabalho. Os valores de AOD_{550} foram obtidos pelo satélite Terra por meio do sensor MODIS (Moderate Resolution Imaging Spectroradiometer) que faz medições sobre a região de Botucatu entre 10:00 UTC e 11:00 UTC e são considerados como valores diários. O erro esperado para AOD_{550} fornecidos pelo MODIS é: $\Delta(AOD_{550}) = 0.05 \pm 0.15(AOD_{550})$ (REMER et al., 2008). Os dados de vapor de água foram obtidos do sensor MODIS a bordo do satélite Terra. O valor de vapor d'água pelo MODIS apresenta erro de 5% a 10% (GAO:KAUFMAN, 2003). Os dados do sensor MODIS foram baixados dos produtos atmosféricos

GIOVANNI

(<http://disc.sci.gsfc.nasa.gov/giovRNAi/overview/index.html>), para uma área de 0,2° de latitude x 0,2° de longitude ao redor do local de medição. Os dados de AOD_{550} e vapor d'água foram separados para condições de céu claro.

7.3.2.2 Turbidez de Linke (TL)

A turbidez de Linke (TL) (LINKE, 1922) expressa o número de atmosferas (secas e limpas) que produz na irradiância direta no topo da atmosfera uma atenuação semelhante a uma atmosfera real (PEDRÓS et al., 1999; HUSSAIN et al., 2000). Modificações e melhorias foram sugeridas para TL (KASTEN, 1996; ZAKY et al., 2004; MAVROMATAKIS:FRANGHIADAKIS, 2007). Neste trabalho, a TL é calculada a partir das metodologias (DJAHER:IRBAH, 2013) e (LI:LAM, 2002), representadas como TL_{Dj} e TL_{Li} , respectivamente.

Metodologia (TL_{Dj}):

A turbidez de Linke (DJAHER:IRBAH, 2013) (TL_{Dj}) é determinada pela equação:

$$TL_{Dj} = T_{lk} \frac{\frac{1}{\delta_{Ra}(m_a)}}{\frac{1}{\delta_{Rk}(m_a)}} \quad (1)$$

Onde: T_{lk} é um fator de correção de TL (eq.2) (TRABELSI:MASMOUDI, 2011; KASTEN, 1980), $\delta_{Rk}(m_a)$ é a espessura óptica integrante de Rayleigh (adimensional)

(eq.3) e $\delta_{Ra}(m_a)$ é a espessura óptica integrante (adimensional) (eq.4) (LOUCHE et al., 1986; KASTEN,1996).

$$T_{lk} = [0.90 + 9.40 \sin(h)] * \left\{ \left[\ln \left(I_0(h) \frac{R_0}{R} \right) \right] - \ln [DNI(h)] \right\} \quad (2)$$

Onde:

h: ângulo de elevação solar (graus).

$I_0(h)$: representa a constante solar ($=1367 \text{ Wm}^{-2}$).

DNI(h): representa a irradiância solar direta na incidência medida (Wm^{-2}).

R_0 e R : são a distância instantânea e a distância média entre a terra-sol, respectivamente. A razão entre R_0 e R denomina-se fator de correção de excentricidade da órbita terrestre (E_0) (Iqbal, 1983).

$$\frac{1}{\delta_{Ra}}(m_a) = 6.6296 + 1.7513m_a - 0.1202m_a^2 + 0.0065m_a^3 - 0.00013m_a^4 \quad (3)$$

$$\frac{1}{\delta_{Rk}}(m_a) = 9.40 + 0.90m_a \quad (4)$$

m_a é massa ótica a pressão real calculada pela equação:

$$m_a = m_r \left(\frac{p}{p_0} \right) \quad (5)$$

m_r é a massa ótica relativa definida em função do ângulo de elevação solar (h) em graus (CAÑADA et al., 1993):

$$m_r = [\sin(h) + 0.15 * (3.885 + h)^{-1.253}]^{-1} \quad (6)$$

A relação p/p_0 é a razão entre a pressão local a pressão padrão (ao nível do mar) e foi calculada em função da altitude (H) local em metros (IQBAL, 1983) pela expressão:

$$\frac{p}{p_0} = \exp(-0.0001184 * H) \quad (7)$$

Metodologia (TL_{Li}):

O cálculo de TL utilizando (LI:LAM, 2002) (TL_{Li}) surge da definição que a DNI para todo espectro solar e sob condição de céu claro pode ser calculado:

$$DNI(h) = I_0(h) \exp(-TL_{Li} \delta_{Rk} m_a) \quad (8)$$

Reorganizado a equação 8 para TL_{Li} :

$$TL_{Li} = \frac{\ln \left(\frac{DNI(h)}{I_0(h)} \right)}{(\delta_{Rk} m_a)} \quad (9)$$

Os parâmetros $\delta_{Rk}(m_a)$, $I_0(h)$ e $DNI(h)$ foram definidos anteriormente. Na determinação de TL_{Li} foi utilizado para o cálculo da massa ótica a pressão real (m_a) (eq.5), a m_r em função do ângulo zenital (Z) em graus (eq. 10):

$$m_r = [\cos Z + 0.15 * (93.885 - Z)^{-1.253}]^{-1} \quad (10)$$

Considerando as equações de TL descritas, os valores de TL_{Dj} e TL_{Li} podem ser facilmente obtidos diretamente das medições de DNI realizadas com o pireliômetro instalado em Botucatu.

Os valores típicos de TL normalmente variam entre 1 e 10, elevados valores significa que DNI é mais atenuada e a atmosfera apresenta alta concentração de aerossóis e grande teor de vapor d'água (atmosfera menos transparente). Neste estudo será considerada a caracterização dos tipos de atmosfera associadas aos níveis de turbidez de Linke (Tabela 1) (LECKNER, 1978). Como a turbidez atmosférica é um índice para condição de céu claro, os critérios adotados na seleção dos dados para a aplicação nos modelos foi baseada em Karayel et al., (1984), onde a DNI deve ser maior que 200 W/m^2 e a razão entre a irradiância difusa pela global menor que 1/3. Foi considerado que a transmissividade atmosférica de R_G (kt) não pode ser inferior a 0.675.

Tabela 1. Tipos de atmosfera para os diferentes níveis de turbidez atmosférica. (LECKER, 1978).

Tipo de atmosfera	TL	Visibilidade (km)
Atmosfera de Rayleigh e pura	1	340
Clara, ar quente	2	28
Turva, umida, ar quente	3	11
Atmosfera poluída	4 - 8	< 5

7.4 RESULTADOS E DISCUSSÃO

7.4.1 Variação diurna e sazonal da turbidez de Linke

Os valores da Turbidez de Linke, com base nas eqs. 1-10, foram calculados para as horas de céu sem nuvens. Os valores médios horários da turbidez com as metodologias (TL_{Dj} e TL_{Li}) são apresentados na Figura 2a. A variação horária de TL_{Dj} é semelhante a TL_{Li} , no entanto, TL_{Li} tem valores em média 10,34 % maiores que

TL_{Dj} . A turbidez tem variabilidade ao longo do dia, aumenta de manhã e reduz à tarde, com máximos valores próximos ao meio dia solar. Os valores máximos coencidem com o período de maior incidência solar no dia. Portanto, nesta análise a variabilidade é explicada com base na elevação da intensidade do fluxo radiante ao longo do dia para cada hora. Outros fatores podem influenciar na oscilação horária da turbidez: as condições meteorológicas, a poluição local, o vapor d'água, os aerossóis originados de incêndios florestais sazonais e as atividades industriais local (BILBAO et al, 2014). No entanto, é importante destacar que a turbidez e a m_a são inversamente proporcionais. Então, de manhã, como a m_a é maior, TL é menor. Esta é uma limitação do método.

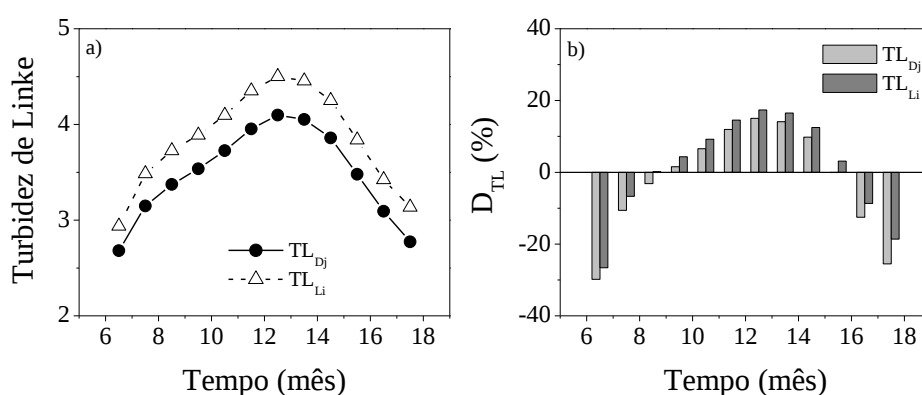


Figura 2. a) Variação horária média da turbidez de Linke e b) desvio relativo para Botucatu entre 1996-2008.

O desvio relativo (D_{TL} , %) foi calculado por meio da relação $D_{TL} (\%) = [(TL_h - TL_m)/TL_h] \cdot 100$, onde TL_m e TL_h são os valores das curvas de TL médio total e médio de cada hora, respectivamente (Figura 2b). Os valores positivos de D_{TL} são indicativos que a média é inferior aos valores horários. Os valores de D_{TL} para TL_{Dj} e TL_{Li} variaram, respectivamente, entre -29,82 e 14,09% e entre -26,58 e 17,35%. Considerando o intervalo de 07h00min às 16h00min, D_{TL} tem menor amplitude e mostra valor média de $D_{TLDj} = 5,02 \pm 8,69\%$ e $D_{TLLi} = 7,89 \pm 8,20\%$. Trabelsi e Masmoudi (2011), destacaram que o vapor d'água na atmosfera, os aerossóis e a variabilidade das diversas variáveis meteorológicas que influenciam na atmosfera são determinantes na amplitude de D_{TL} .

A atenuação de DNI no período diurno nas estações do ano foi apresentada como valores médios para manhã (até as 09h30min), próximos ao meio dia solar (entre 10h30min e 13h30min) e para tarde (depois das 13h30min) (Figura 3a, b).

Para esta análise foram utilizadas todas as horas de céu claro do período estudado. O maior grau de transmissão de DNI é observado na parte da manhã para a primavera e verão. Próximo ao meio dia solar a transmissão é menor, e consequentemente, maior o valor da turbidez atmosférica. Isto está relacionado ao estado óptico da atmosfera nas estações do ano. A turbidez atmosférica possui variação diferente no período quente/chuvoso (primavera e verão) e frio/seco (outono e inverno), mas a variação de TL_{Dj} (Figura 3a) é similar a TL_{Li} (Figura 3b). A variabilidade diurna da turbidez atmosférica é mais evidente no período de aquecimento, enquanto que no período frio o grau de turbidez é pouco variado (USCKA-KOWALKOWSKA, 2013).

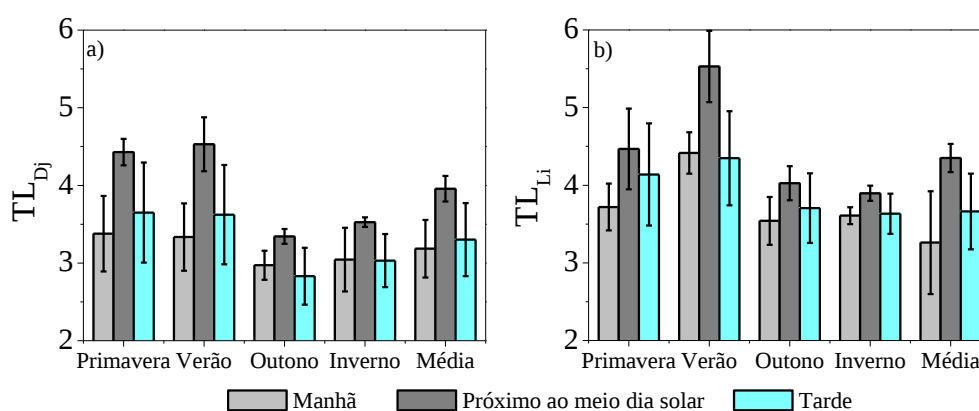


Figura 3. Variabilidade sazonal da turbidez de Linke em Botucatu para os anos de 1996-2008. a) metodologia TL_{Dj} e b) metodologia TL_{Li} .

Entre as estações do ano, a turbidez atmosférica possui variabilidade média de $\approx 17\%$. Na evolução diurna, o valor médio máximo é observado no verão ($TL_{Dj} = 3.83 \pm 0.62$) e o médio mínimo no outono ($TL_{Dj} = 3.05 \pm 0.26$). A menor transparência atmosférica (maior turbidez) ocorre na primavera e verão. O valor do TL_{Dj} para a primavera próximo ao meio dia solar é (4.43 ± 0.17) e para o verão no período de manhã e a tarde são 3.33 ± 0.43 e 3.62 ± 0.64 , respectivamente (Tabela 2). Na evolução diurna, a maior transparência (menor turbidez), ocorre no outono e inverno. O valor de TL_{Li} para o outono de manhã e a tarde é inferior a 4. Quando comparado com os valores de manhã e da tarde os maiores TL são observados em torno da culminação solar.

Tabela 2. Variabilidade sazonal da turbidez de Linke em Botucatu para os anos de 1996-2008.

	Estação	Manhã	Próximo ao meio dia solar	Tarde
TL _{Dj}	Primavera	3,38±0,49	4,43±0,17	3,65±0,64
	Verão	3,33±0,43	4,53±0,35	3,62±0,64
	Outono	2,97±0,19	3,34±0,09	2,83±0,36
	Inverno	3,04±0,41	3,52±0,06	3,03±0,34
	Média	3,18±0,37	3,95±0,16	3,30±0,47
TL _{Li}	Primavera	3,72±0,30	4,47±0,52	4,14±0,66
	Verão	4,41±0,26	5,53±0,46	4,35±0,60
	Outono	3,54±0,31	4,02±0,22	3,70±0,45
	Inverno	3,61±0,11	3,89±0,10	3,63±0,26
	Média	3,26±0,66	4,35±0,20	3,66±0,49

7.4.2 Análise da turbidez de Linke média mensal

Os valores da turbidez média mensal foram obtidos dos valores horários da turbidez correspondentes a cada mês. A evolução anual média mensal da turbidez de Linke e os valores médios por estação do ano, para TL_{Dj} e TL_{Li}, são apresentados na Figura 4 (a, b). Os valores da turbidez atmosférica indicam que maio e julho foram os meses de menor atenuação de DNI (maior transparência atmosférica), enquanto que janeiro foi o mês de menor transparência (maior atenuação de DNI) (Figura 4a).

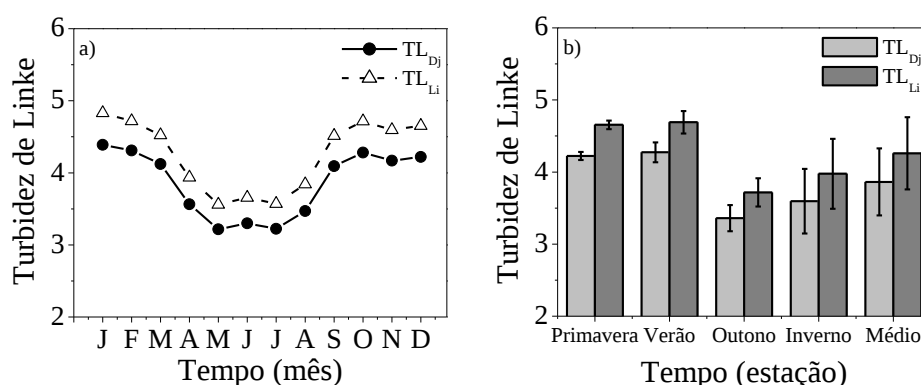


Figura 4. Variação da turbidez de Linke em Botucatu, nos anos de 1996-2008. a) Variação Média anual. B) Variação média anual por estação.

O aumento da turbidez atmosférica também é visível nos demais meses do verão e da primavera. Entre os meses de abril a agosto a atmosfera é menos turva (TL_{Dj} menores que 3,46 e TL_{Li} menores que 3,93). A elevação da intensidade do

fluxo radiante ao longo do dia tem grande influência na alteração da turbidez atmosférica em diferentes escalas, mas o vapor d'água e os aerossóis são os mais determinantes na variabilidade sazonal.

O período de maior e menor transparência atmosférica é evidenciado quando se considera as estações do ano (Figura 4b). A maior atenuação de DNI ocorre na primavera ($TL_{Dj} = 4.22 \pm 0.05$ e $TL_{Li} = 4.65 \pm 0.06$) e no verão ($TL_{Dj} = 4.27 \pm 0.14$ e $TL_{Li} = 4.69 \pm 0.15$), meses nos quais o deslocamento de materiais particulados que avançam para o estado de São Paulo com os Sistemas Frontais do Atlântico Sul (durante a primavera, principalmente) e com a Zona de Convergência do Atlântico Sul (ZCAS) (no verão) são as mais elevadas no ano (ALLEN et al., 2004). Na primavera e no verão o teor de vapor d'água na atmosfera e o aumento de aerossóis (Figura 5a, b), com origem da queima de biomassa (cana de açúcar, pastagem, florestas e matas nativas) para produção agrícola e agropecuária em regiões próximas ao Estado de São Paulo e da região amazônica (HOLBEN et al., 2001, FRANÇA et al., 2014), são os principais fatores responsáveis pela elevação da turbidez atmosférica e atenuação de DNI em Botucatu.

No verão o aumento da turbidez é causado, principalmente, pela elevação da concentração de vapor d'água na atmosfera. O aumento da turbidez na atmosfera devido o vapor d'água e aos aerossóis também estão ligados à circulação atmosférica, que é importante no transporte de massas de ar e na variação temporal da turbidez (ELMINIR et al., 2006, FREITAS et al., 2005). Na estação fria (outono e inverno), dias com precipitações reduzidos, a atmosfera é frequentemente mais clara com menor concentração de vapor d'água e aerossóis na atmosfera local. A variabilidade sazonal da turbidez atmosférica descrita na literatura (GRENIER et al., 1995; POWER:GOYAL, 2003; USCKA-KOWALKOWSKA, 2013), assim como a observada em Botucatu/SP, mostram os valores máximos na primavera e no verão e mínimo no outono e inverno.

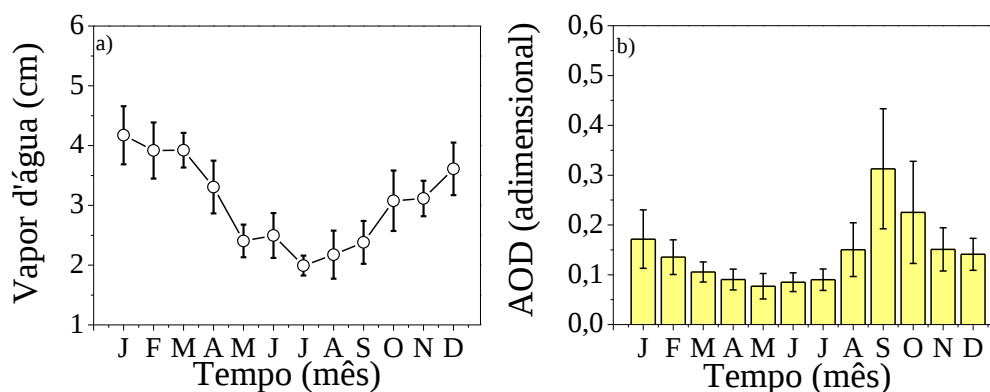


Figura 5. Médias mensais do período 2000 a 2008, considerando apenas dias de céu limpo. a) Vapor d'água e b) AOD.

Os valores médios mensais anual da turbidez mostram baixa variação entre os 13 anos analisados (Tabela 3), com TL_{Dj} máximo ($5,70 \pm 1,18$) em outubro de 2001 e valor mínimo de ($2,84 \pm 0,51$) em agosto de 1997. O TL_{Li} tem mínimo ($3,15 \pm 0,57$) em agosto de 1997 e máximo ($6,30 \pm 1,31$) em outubro de 2001. Os valores de TL_{Li} foram 9,31% maiores que TL_{Dj} , o mais provável é que esta diferença esteja relacionada com as equações utilizadas. A turbidez indicou tendência crescente e este efeito pode ser resultante do aumento das queimadas (desmatamento ilegal na floresta amazônica e incêndios diversos), da crescente urbanização e expansão industrial local e regional. Dependendo do local, a urbanização é principal fator responsável pelo aumento da poluição e da turbidez atmosférica em grandes regiões (Rahoma e Hassan, 2012). Os índices de turbidez obtidos para Botucatu são similares aos encontrados em 3 regiões de Bangladesh, TL entre 3,46-4,83 (HUSSAIN et al., 2000) e em locais na África (TL = 3,50) (DIABATÉ et al., 2003). A turbidez atmosférica com TL_{Dj} e TL_{Li} são menores que os obtidos em Hong Kong, TL entre 3,70-5,26 (LI:LAM, 2002). Apesar desses resultados, cada local possui características e condições atmosféricas distintas, consequentemente, diferentes níveis de turbidez são encontrados.

Tabela 3. Turbidez atmosférica de Linke média mensal com TL_{Dj} e TL_{Li} para Botucatu, 1996-2008.

Turbidez de Linke (TL _{Dj})													
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
J	--	3,90±0,72	4,46±0,67	4,78±0,90	4,43±0,85	4,30±0,89	4,14±0,90	4,25±1,06	4,55±1,52	4,43±1,20	4,10±1,33	4,99±1,45	4,47±1,91
F	4,50±0,77	3,94±0,63	4,42±0,87	4,14±0,93	4,19±0,91	4,82±1,43	4,02±1,36	3,95±0,78	4,26±1,07	4,05±0,86	4,23±0,99	3,92±0,81	5,71±2,07
M	4,48±1,10	3,90±0,68	4,22±0,85	3,82±0,93	4,36±0,98	4,04±0,69	3,67±0,59	4,55±1,59	3,94±0,70	3,94±0,76	4,15±0,85	4,37±1,74	4,35±1,27
A	3,64±0,61	3,45±0,73	3,41±0,86	3,44±0,62	3,21±0,51	3,50±0,84	3,44±0,51	4,16±2,18	3,92±1,20	4,05±0,76	3,06±0,49	3,80±0,75	3,57±0,98
M	3,33±0,53	3,17±0,61	2,98±0,56	3,14±0,79	3,31±0,77	3,32±1,16	3,23±1,11	2,99±0,63	3,16±0,59	3,56±1,04	2,98±0,50	3,57±1,11	3,19±0,61
J	3,56±1,49	3,03±0,76	3,28±0,80	2,97±0,51	3,60±0,81	3,78±1,23	3,10±0,51	3,38±1,07	3,51±1,13	3,33±0,84	3,04±0,60	3,09±0,71	3,21±0,66
J	3,21±0,63	3,41±0,73	3,58±0,73	3,36±0,82	3,09±0,75	3,34±0,91	2,96±0,77	3,18±0,52	3,32±0,78	3,06±0,69	3,26±0,70	3,01±0,64	3,03±0,76
A	3,50±0,99	2,84±0,51	4,57±0,72	3,60±0,83	3,44±0,72	3,55±0,83	3,73±0,89	3,27±0,86	3,18±0,98	3,64±1,04	3,53±0,78	3,11±0,64	3,31±0,64
S	4,03±1,00	4,20±0,85	4,59±1,16	4,30±0,88	3,98±0,42	4,20±1,33	4,48±1,03	3,64±0,85	4,47±0,92	4,28±1,39	3,64±0,69	3,88±0,81	3,84±1,01
O	4,09±0,67	3,04±0,00	4,12±0,83	4,24±0,96	4,52±0,97	4,54±1,86	4,74±0,59	3,77±0,99	3,69±0,86	4,53±0,74	4,20±0,93	4,36±0,70	4,40±0,69
N	3,56±0,66	--	3,90±0,57	4,00±0,81	4,34±1,12	5,70±1,18	4,17±0,96	4,18±1,10	4,01±0,84	3,90±0,75	4,12±0,96	3,89±0,99	4,10±0,65
D	4,56±0,82	--	4,31±0,69	4,45±0,99	4,51±0,90	3,88±1,35	3,79±0,71	4,39±0,77	4,43±0,90	4,20±0,92	4,13±0,68	4,22±1,37	3,87±0,85
Média	3,86±0,84	3,49±0,62	3,99±0,78	3,85±0,83	3,91±0,81	4,08±1,14	3,79±0,83	3,81±1,03	3,87±0,96	3,92±0,92	3,70±0,79	3,85±0,98	3,92±1,01
Turbidez de Linke (TL _{Li})													
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
J	--	4,32±0,79	4,94±0,74	5,29±0,99	4,90±0,94	4,76±0,98	4,58±0,99	4,70±1,17	4,99±1,59	4,72±1,50	4,54±1,48	5,51±1,60	4,69±2,07
F	4,84±0,83	4,36±0,70	4,88±0,96	4,58±1,02	4,64±1,01	5,24±1,39	4,46±1,50	4,37±0,86	4,71±1,18	4,48±0,95	4,68±1,09	4,34±0,89	5,86±1,81
M	4,90±1,20	4,31±0,74	4,67±0,94	4,23±1,03	4,83±1,08	4,47±0,76	4,06±0,65	5,02±1,76	4,36±0,77	4,36±0,83	4,47±0,69	4,83±1,91	4,55±1,04
A	4,08±0,66	3,79±0,71	3,77±0,97	3,81±0,69	3,55±0,57	3,88±0,93	3,81±0,57	4,50±2,30	4,34±1,33	4,49±0,85	3,39±0,54	4,20±0,83	3,95±1,09
M	3,83±0,59	3,52±0,67	3,30±0,62	3,47±0,89	3,63±0,89	3,67±1,30	3,57±1,24	3,25±0,74	3,49±0,65	3,95±1,15	3,30±0,56	3,96±1,23	3,49±0,69
J	4,10±1,58	3,34±0,84	3,64±0,88	3,29±0,56	3,97±0,93	4,17±1,37	3,44±0,57	3,75±1,19	3,84±1,30	3,69±0,93	3,35±0,69	3,37±0,82	3,55±0,75
J	3,73±0,71	3,78±0,81	3,95±0,83	3,72±0,91	3,30±0,83	3,62±0,83	3,26±0,86	3,52±0,57	3,68±0,86	3,40±0,76	3,61±0,77	3,34±0,71	3,35±0,84
A	4,00±1,11	3,15±0,57	5,06±0,79	3,99±0,92	3,81±0,80	3,93±0,91	4,14±0,98	3,59±0,99	3,52±1,09	4,03±1,16	3,88±0,88	3,44±0,71	3,58±0,62
S	4,50±1,09	4,65±0,94	5,08±1,28	4,76±0,97	4,36±0,50	4,65±1,48	4,87±1,09	4,02±0,95	4,83±0,84	4,72±1,55	4,03±0,77	4,29±0,89	4,25±1,11
O	4,46±0,74	3,36±0,00	4,56±0,92	4,69±1,06	4,89±0,96	5,01±2,06	5,24±0,65	4,17±1,09	4,09±0,96	5,02±0,81	4,61±0,98	4,82±0,77	4,87±0,76
N	3,77±0,72	--	4,31±0,63	4,43±0,89	4,80±1,23	6,30±1,31	4,61±1,06	4,62±1,21	4,44±0,92	4,26±0,90	4,55±1,05	4,30±1,09	4,54±0,72
D	4,83±0,89	--	4,77±0,77	4,92±1,09	4,99±0,99	4,30±1,49	4,20±0,78	4,86±0,85	4,88±1,01	4,65±1,01	4,57±0,75	4,67±1,51	4,28±0,94
Média	4,28±0,92	3,86±0,68	4,41±0,86	4,26±0,92	4,31±0,89	4,50±1,23	4,19±0,91	4,20±1,14	4,27±1,04	4,31±1,03	4,08±0,85	4,26±1,08	4,25±1,04

7.4.3 Comparação de TL calculada em Botucatu com as disponíveis na Website SoDA

Inserindo a latitude, longitude e a altitude do local de interesse é possível obter a turbidez de Linke mensal por mapas globais para qualquer lugar do mundo através da website SoDA (www.soda-is.com). A turbidez de Linke (TL_{SoDA}) é baseada numa série de informações global de satélites, vapor d'água, aerossóis e a partir de informações sobre turbidez na superfície (HOVE:MANYUMBU, 2013). Os valores de TL mensais (Tabela 4) disponíveis em SoDA são maiores na primavera/verão e menores no outono/inverno, esta sazonalidade se assemelha as obtidas neste trabalho com TL_{Dj} e TL_{Li} . Concordando com a maior transparência atmosférica no outono/inverno e menor na primavera/verão. A turbidez atmosférica calculada através de TL_{Dj} e TL_{Li} é, respectivamente, 15,80% e 23,52% superior que TL_{SoDA} . O coeficiente de variação (cv) expresso em porcentagem é um indicativo da dispersão dos dados, foi utilizado neste estudo. TL_{Dj} tem $cv = 12,04\%$, TL_{Li} com $cv = 11,76\%$ e TL_{SoDA} com coeficiente de variação igual a 9,66%. Estimativas com satélites estão sujeitas a maiores erros, então os resultados obtidos neste trabalho são mais representativos, pois foram determinados com medidas em superfície. No entanto, não podemos desconsiderar as estimativas fornecidas por SoDA.

Tabela 4. Turbidez de Linke disponível em SoDA (www.soda-is.com).

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
TL_{SoDA}	3,40	3,70	3,70	3,60	3,30	3,10	2,80	2,80	3,40	3,00	3,20	3,10

As alterações mensais podem ser verificadas considerando a modulação do perfil observado. A amplitude de variação da turbidez é calculada por: $[TL(max)-TL(min)]/[TL(max)+TL(min)]$. Com base nos dados locais, a modulação com TL_{Dj} e TL_{Li} foram respectivamente, 15,41 e 15,14%. Utilizando os resultados de TL_{SoDA} a modulação é de 13,84%. Na ilha de Creta a modulação encontrada foi ~18% (MAVROMATAKIS:FRANGHIADAKIS, 2007).

7.4.4 Correlações polinomiais para turbidez de Linke

Da Figura 5 foram ajustados modelos polinomiais do quarto grau (Figura 6) para a variação temporal de TL_{Dj} e TL_{Li} , respectivamente, (Eqs. 11 e 12) e mostra como ocorre a variabilidade para cada mês ao longo do ano. Este grau polinomial foi escolhido por apresentar elevado coeficiente de correlação e menor erro. Os coeficientes dependem do mês do ano como entrada. A aplicação principal desta regressão polinomial é no preenchimento de algumas lacunas na série de dados locais ou em regiões com condições climáticas semelhantes.

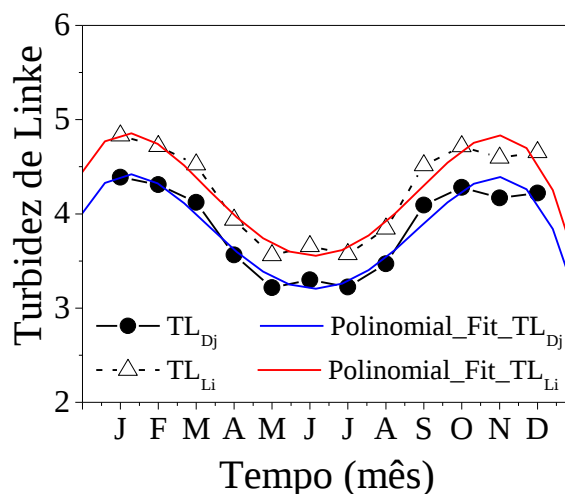


Figura 6. Regressão polinomial da curva da variação da turbidez de Linke em Botucatu.

$$TL_{Dj} = 4,03 + 0,71T - 0,38T^2 + 0,05T^3 - 0,002T^4 \quad (11)$$

$$TL_{Li} = 4,47 + 0,72T - 0,39T^2 + 0,06T^3 - 0,002T^4 \quad (12)$$

Os coeficientes de determinação foram elevados e iguais ($R^2 = 0,88$). Este resultado mostra que existe boa relação entre a variação da turbidez mensal e os meses do ano em Botucatu. Por isso, é possível obter o valor da turbidez atmosférica média mensal. Em 28 locais da Itália (CUCUMO et al., 2000) geraram equações polinomiais do oitavo grau com erro de $\pm 5\%$. O resultado local assemelha-se ao encontrado para a região de Ghärdaia –

província da Argélia onde ajustaram equações polinomiais do 6º grau em função dos anos (DJAHER:IRBAH, 2013).

7.4.5 Relação entre a turbidez de Linke e os parâmetros meteorológicos

A turbidez atmosférica depende da variação de curto prazo das condições meteorológicas locais (temperatura do ar, velocidade e direção do vento) e a longo prazo da variabilidade do clima da região (CHAÛBANE et al., 2004). O vento (velocidade e direção), por exemplo, pode transportar bastante umidade e materiais particulados suspensos no ar (aerossóis) a partir de fontes primitivas distantes, e desempenham um papel importante na variação espacial e temporal da turbidez atmosférica (MASMOUDI et al, 2003). Por isso, a influência da velocidade do vento média com o turbidez de Linke em Botucatu é analisado.

A turbidez atmosférica em Botucatu cresce com o aumento da velocidade do vento (Figura 7a). Intervalos de 0,40 m/s foram separados para a velocidade do vento visando melhor análise da variabilidade temporal. Neste estudo a velocidade do vento ficou compreendida no intervalo de 3,40 e 7,40 m/s, com valores máximos predominantes na primavera. Nos valores de velocidade do vento menor que 5,0 m/s, a turbidez média é $TL_{Dj} = 3,92 \pm 0,54$ e $TL_{Li} = 4,32 \pm 0,58$. Quando a velocidade do vento atinge média de 7,40 m/s, TL pode alcançar valores médios iguais a $TL_{Dj} = 4,51 \pm 0,58$ e $TL_{Li} = 4,97 \pm 0,27$. Estes resultados são característicos dos efeitos locais, tanto das condições climáticas quanto da influência humana.

A influência da velocidade do vento foi analisada em função do intervalo de variação da turbidez. Com a velocidade do vento $\geq 5,0$ m/s foi encontrado 8,59% da turbidez com valores de $TL_{Dj} \geq 4,50$. Quando $TL_{Li} < 5,0$, 73,54% das observações são para velocidades do vento entre 3,0 e 5,0m/s. Nos resultados locais a velocidade do vento correlaciona positivamente com a turbidez atmosférica com $r = 0,355$ para TL_{Dj} e $r = 0,353$ com TL_{Li} . A turbidez atmosférica pode correlacionar positivamente ou negativamente com o vento. No Cairo – Egito, por exemplo, a turbidez atmosférica diminui com o aumento da velocidade do vento, Elminir et al., (2006), que pode ser explicado pelo fluxo de ar causado pelas construções.

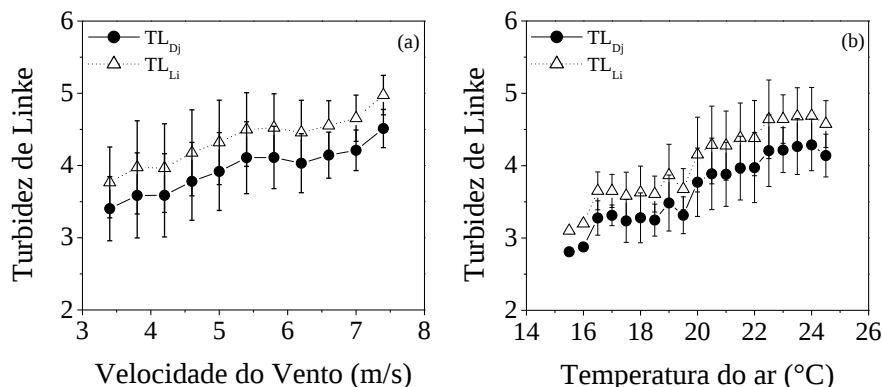


Figura 7. Variação de TL_{Dj} e TL_{Li} em função: (a) Velocidade do vento, (b) Temperatura do ar.

A Figura 7b apresenta a variação da turbidez de Linke TL_{Dj} e TL_{Li} em função da temperatura do ar. Semelhante a velocidade do vento, a temperatura do ar é de extrema importância em estudos climáticos, pois varia no espaço, no tempo e em função da altitude. Neste sentido, é possível analisar a variabilidade da turbidez com a temperatura do ar variando no intervalo de $[15,50-24,50^{\circ}\text{C}]$, em intervalo de $0,50^{\circ}\text{C}$. A TL cresce linearmente com a temperatura do ar. O valor mínimo da turbidez ($TL_{Dj} = 2,81 \pm 0$ e $TL_{Li} = 3,10 \pm 0$) ocorre quando a temperatura média do ar é $15,50^{\circ}\text{C}$ e o máximo ($TL_{Dj} = 4,14 \pm 0,29$ e $TL_{Li} = 4,57 \pm 0,32$) quando a temperatura do ar é $24,50^{\circ}\text{C}$. Os menores valores ocorrem essencialmente no inverno (baixas temperaturas e atmosfera seca).

Quando a temperatura média do ar é igual a $19,50^{\circ}\text{C}$, observa-se o decaimento da turbidez, isto associa-se a frequência de dias no outono e inverno, onde a atmosfera estava clara com predominância de valores da turbidez de Linke menor que 3,50. A temperatura do ar correlaciona positivamente com a turbidez atmosférica com $r = 0,71$ para TL_{Dj} e $r = 0,70$ com TL_{Li} . Estes resultados confirmam a influência da velocidade do vento e da temperatura do ar na variabilidade temporal da turbidez atmosférica no local estudado.

7.4.6 Distribuição de frequência da turbidez de Linke

A distribuição de frequência na classificação da turbidez atmosférica é um indicativo importante na caracterização das classes predominantes nos intervalos de

variações considerados. A frequência de ocorrência e a frequência acumulada de TL_{Dj} e TL_{Li} são observados em (Figura 8a, b), respectivamente. A distribuição de frequência foi separada no intervalo 1,95 e 12,00 com intervalo de 0,30. Foram utilizadas 20.124 horas de céu claro. A Figura 8 (a) mostra uma distribuição modal, com máximo em 2,85 com TL_{Dj} e 3,15 para TL_{Li} , representando 16,07% e 14,59% das observações, respectivamente. O menor valor é observado para TL igual a 12. Para TL_{Li} cerca de 92,06% das ocorrências encontram-se no intervalo entre [2,0-6,0] e 93,95% das ocorrências deste intervalo para TL_{Dj} . Analisando quatro metodologias para o cálculo de TL, Wen e Yeh (2009) observaram diferentes variações com: [0,5-7,0] metodologia 1; [2,0-10,0] metodologia 2; [4,0-10] metodologia 3; [5,0-10,0] metodologia 4, com valores inclinados para a esquerda e pico entre 5 e 8.

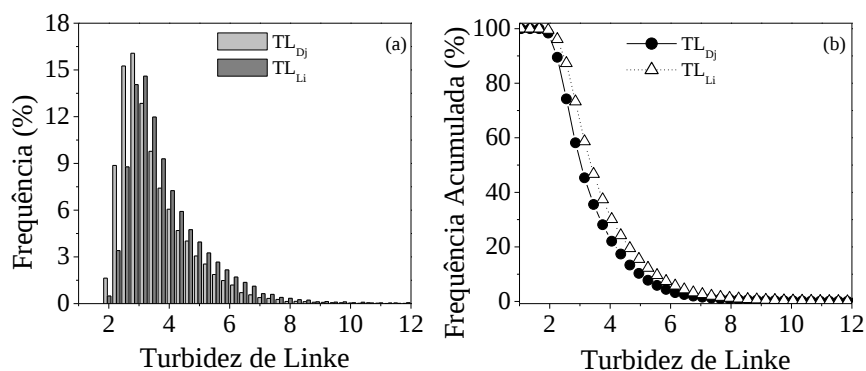


Figura 8. Turbidez atmosférica TL_{Dj} e TL_{Li} entre os anos 1996 e 2008. (a) Distribuição de frequência horária e (b) Frequência acumulada horária.

A frequência acumulada de distribuição da turbidez horária (Figura 8b) serve como indicativo da porcentagem de observações (instantâneas de horas de céu claro) em que um dado nível de turbidez é atingido. Os resultados obtidos mostram que quando $TL_{Dj} \leq 2,0$ ocorreram $\approx 41,83\%$ horas de céu limpo, representando uma atmosfera de Rayleigh pura, clara e ar quente. Um valor de $\approx 30,50\%$ das horas de céu claro estão entre $(2,0 < TL_{Dj} \leq 4,0)$, que mostra uma atmosfera turva, úmida e ar quente. Cerca de $\approx 28,12\%$ das horas, a turbidez calculada excedeu 4,0, representando uma atmosfera poluída. Quando comparada com TL_{Dj} a turbidez com TL_{Li} , tem uma redução de 15,10% na condição de atmosfera de Rayleigh pura,

clara e ar quente; aumento de 5,82% e 9,28% na condição de atmosfera turva, úmida e ar quente; e poluída, respectivamente.

7.5 CONCLUSÕES

O presente estudo calcula a extinção de DNI na região de Botucatu, para valores horários e médios mensais. A extinção de DNI é dada por meia da turbidez atmosférica de Linke. As metodologias TL_{Dj} e TL_{Li} foram adotadas. A turbidez tem variabilidade ao longo do dia, aumenta de manhã e reduz à tarde, com máximos valores próximos ao meio dia solar. A turbidez horária com TL_{Li} calcula valores 10,34% maiores que TL_{Dj} . A variabilidade média horária diurna de TL é intensificada pela elevação da intensidade do fluxo radiante. Para a variabilidade diária anual, na primavera e no verão o vapor d'água na atmosfera e os aerossóis são os principais fatores responsáveis pela elevação da turbidez atmosférica. No verão o aumento da turbidez é causado, principalmente, pela elevação do vapor d'água na atmosfera.

Quando relacionado com as variáveis meteorológicas, a turbidez cresce com o aumento da velocidade do vento e da temperatura do ar, ou seja, correlacionam positivamente com a turbidez atmosférica. As maiores ocorrências de atmosfera turva e ar quente foram na primavera e no verão, período com menor transparência atmosférica. Aproximadamente $\approx 28,12\%$ das horas analisadas, a turbidez obtida excedeu o valor de 4,0, representando uma atmosfera poluída.

O aumento das queimadas, a crescente urbanização e a expansão industrial local e regional são responsáveis pela tendência crescente da turbidez. No Brasil as queimadas de matas nativas para a produção agrícola e pecuária na região da floresta amazônica, assim como as queimadas agrícolas e a queima de cana de açúcar em regiões adjacentes a Botucatu são grandes vilões no aumento de materiais particulados na atmosfera. Na região Sudeste do Brasil, devido à circulação atmosférica, estes acontecimentos alteram o padrão da atmosfera. No verão, a circulação atmosférica também é responsável por transportar bastante umidade da região amazônica para a região sudeste do Brasil. Portanto, a atmosfera de Botucatu mostra ser influenciada pelos aerossóis e vapor d'água de origens diferentes.

Estes dois parâmetros alteram a distribuição de DNI na atmosfera e são determinantes na densidade do fluxo radiante na superfície.

Há uma lei estadual em São Paulo que institui uma redução na prática da queima de cana de açúcar, no entanto, o resíduo sólido da cana moída (bagaço da cana) é usado como combustível industrial renovável, por isso as emissões de combustão de biomassa é provável que não tenha reduzido (Scaramboni et al., 2015). Este trabalho mostra que o efeito no período de estudo é significativo e política publica ambiental devem ser consideradas.

Portanto, com o desmatamento de florestas nativas e a crescente urbanização, novos trabalhos devem ser realizados na perspectiva de analisar os novos padrões na atmosfera e na distribuição de DNI. Novas pesquisas tornam-se indispensáveis, pois estudos recentes mostram que queimadas acidentais e intencionais causados pelo homem têm aumento. Estes incêndios acontecem, principalmente, nos estados da região da floresta amazônica e causam poluição atmosférica com prejuízos a saúde de milhões de pessoas e a aviação.

7.6 REFERÊNCIAS BIBLIOGRÁFICAS

- ALLEN, A. G.; CARDOSO, A. A.; DA ROCHA, G. O. Influence of sugar cane burning on aerosol soluble ion composition in Southeastern Brazil. **Atmospheric Environment**, v. 38, p. 5025 – 5038, 2004.
- BATLLES, F. J.; RUBIO, M. A.; TOVAR, J.; OLMO, F. J.; ALADOS-ARBOLEDAS, L. Empirical modeling of hourly direct irradiance by means of hourly global irradiance. **Energy**, v.25, p.675-688, 2000.
- BILBAO, J.; ROMÁN, R. MIGUEL, A. Turbidity coefficients from normal direct solar irradiance in central Spain. **Atmospheric Research**, v. 143, p.73-84, 2014.
- CAÑADA, J.; PINAZO, J. M.; BOSCA, J. V. Determination of Angström's turbidity coefficient at Valencia. *Renew. Energy*, v.3, p.621–626, 1993.
- CHAËBANE, M.; MASMOUDI, M.; MEDHIOUB, K. Determination of Linke turbidity factor from solar radiation measurement in northern Tunisia. **Renewable Energy**, v.29, p.2065–2076, 2004.

- CODATO, G.; OLIVEIRA, A. P.; SOARES, J.; ESCOBEDO, J. F.; GOMES, E. G.; DAL PAI, A. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. **Theoretical and Applied Climatology**, v. 93, p.57–73, 2008.
- CUCUMO, M.; KALIAKATSOS, D.; MARINELLI, V. A calculation method for the estimation of the Linke turbidity factor. **Renewable Energy**, v.19, p.249–58, 2000.
- DAL PAI, A.; ESCOBEDO, J. F.; DAL PAI, E.; SANTOS, C. M. Estimation of Hourly, Daily and Monthly Mean Diffuse Radiation Based on MEO Shadowring Correction. **Energy Procedia**, v.57, p.1150–1159, 2014.
- DIABATÉ, L.; REMUND, J.; WALD, L. Linke turbidity factors for several sites in Africa. **Solar Energy**, 75, 111–119, 2003.
- DJAFER, D.; IRBAH, A. Estimation of atmospheric turbidity over Ghardaïa city. **Atmospheric Research**, v.128, p.76–84, 2013.
- DOS SANTOS, C. M.; SOUZA, J. L.; FERREIRA JUNIOR, R A.; TIBA, C.; MELO, R. O.; LYRA, G. B.; TEODORO, I.; LYRA, G. B.; LEMES, M. A. M. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. **Energy**, v.71, p.388–398, 2014.
- ELLOUZ, F.; MASMOUDI, M.; MEDHIOUB, K. Study of the atmospheric turbidity over Northern Tunisia. **Renewable Energy**, v.51, p. 513–517, 2013.
- ELMINIR, K. H.; HAMID, R. H.; EL-HUSSAINY, F.; AHMED, E.; GHITAS BEHEARY, M. M.; KHALED, M.; ABDEL-MONEIM. The relative influence of the anthropogenic air pollutants on the atmospheric turbidity factors measured at an urban monitoring station. **Science of the Total Environment**, v.368, p.732–743, 2006.
- ELTBAAKH, Y. A.; RUSLAN, M. H.; ALGHOUL, M. A.; OTHMAN, M. Y.; SOPIAN, K. Issues concerning atmospheric turbidity indices. **Renewable and Sustainable Energy Reviews**, v.16, p. 285–6294, 2012.
- ESCOBEDO, J. F.; GOMES, E. N.; OLIVEIRA, A. P.; SOARES, J. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. **Renewable Energy**, v.36, p.169-178, 2011.
- FLORES, J. L.; KARAM, H. A.; FILHO, E. P. M.; FILHO, A. J. P. Estimation of atmospheric turbidity and surface radiative parameters using broadband clear sky solar irradiance models in

Rio de Janeiro-Brasil. **Theoretical and Applied Climatology**, 2015. DOI 10.1007/s00704-014-1369-7.

FORMENTI, P.; WINKLER, H.; FOURIE, P.; PIKETH, S.; MAKGOPA, B.; HELAS, G.; ANDREA, M. O. Aerosol optical depth over a remote semi-arid region of South Africa from spectral measurements of the daytime solar extinction and the nighttime stellar extinction. **Atmospheric Research**, v.62, p.11–32, 2002.

FRANÇA, D.; LONGO, K.; RUDORFF, B.; AGUIAR, D.; FREITAS, S.; STOCKLER, R.; PEREIRA, G. Pre-harvest sugarcane burning emission inventories based on remote sensing data in the state of São Paulo, Brazil. **Atmospheric Environment**, v.99, p. 446-456, 2014.

FREITAS, S. R.; LONGO, K. M.; DIAS, M. A. F. S.; DIAS, P. L. S.; CHATFIELD, R.; PRINS, E.; ARTAXO, P.; GRELL, G. A.; RECUERO, F. S. Monitoring the Transport of Biomass Burning Emissions in South America. **Environmental Fluid Mechanics**, v.5, p.135–167, 2005.

GAO, B. C.; KAUFMAN, Y. J. Water vapour retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels. **Journal Geophysical Research**, 108 (D13), 4389, 2003. <http://dx.doi.org/10.1029/2002JD003023>.

Grenier, J. C.; de La Casinière, A.; Cabot, T. Atmospheric turbidity analyzed by means of standardized Linke's turbidity factor. **J. Appl. Meteorol.** 34, p. 1449–1458, 1995.

GUEYMARD, C. A. Temporal variability in direct and global irradiance at various time scales as affected by aerosols. **Solar Energy**, v.86, p.3544–3553, 2012.

HOLBEN, B. N. ET AL. An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. **Journal of Geophysical Research**, v. 106, n. 11, p. 12.067–12.097, 2001.

HOVE, T.; MANYUMBU, E. Estimates of the Linke turbidity factor over Zimbabwe using ground-measured clear-sky global solar radiation and sunshine records based on a modified ESRA clear-sky model approach. **Renewable Energy**, v.52, p.190-196, 2013.

HUSSAIN, M.; KHATUN, S.; RASUL, M.G. Determination of atmospheric turbidity in Bangladesh. **Renewable Energy**, v. 20, p.325–332, 2000.

IBGE. **Diretoria de Pesquisas - DPE - Coordenação de População e Indicadores Sociais - COPIS**. 2014. Acesso em: <10/12/2015>. Disponível em: < <http://cod.ibge.gov.br/2378L>>.

IQBAL, M. **An introduction to solar radiation**. New York: Academic Press, 1983, 390p.

- JONES, C.; CARVALHO, L. M. V.; HIGGINS, R. W.; WALISER, D. E.; SCHEMM, J. K. E. Climatology of Tropical Intraseasonal Convective Anomalies: 1979 – 2002. **Journal of Climate**, v. 17, p. 523-539, 2004.
- KARAYEL, M.; NAVVAB, M. NE'EMAN, E.; SELKOWITZ, S. Zenith luminance and sky luminance distributions for daylighting calculations. **Energy and Buildings**, v. 6, n.3, p. 283-91. 1984.
- KASTEN, F. A simple parameterization of the pyrhelimeter formula of the Linke turbidity factor. **Meteor Rundsch**, v.33, p.124–127, 1980.
- KASTEN, F. The Linke turbidity factor based on improved values of the integral Rayleigh optical thickness. **Solar Energy**, v.56, n.3, p.239-44, 1996.
- LECKNER, B. The spectral distribution of solar radiation at the earth's surface-elements of a model. **Solar Energy**, v. 20, p. 143-150, 1978.
- LI, D. H. W.; LAM, J. C. A study of atmospheric turbidity for Hong Kong. **Renewable Energy**, v.25, p. 1-13, 2002.
- LINKE, E. Transmission-koeffizient und Trübungsfaktor. **Beiträge zur Physik der freien Atmosphäre**, v. 10, p.91-103, 1922.
- LÓPEZ, G.; BATLLES, F. J. Estimate of the atmospheric turbidity from three broad-band solar radiation algorithms: A comparative study. **Annales Geophysicae**, v.22, p.2657–2668, 2004.
- LOUCHE, A.; PERI, G.; IQBAL, M. An analysis of Linke turbidity factor. **Solar Energy**, v.37, n.6, p.393-6, 1986.
- MASMOUDI, M.; CHAËBANE, M.; MEDHIOUB, K.; ELLEUCH, F. Variability of aerosol optical thickness and atmospheric turbidity in Tunisia. **Atmosphere Research**, v. 66, p.175–88, 2003.
- MAVROMATAKIS, F.; FRANGHIADAKIS, Y. Direct and indirect determination of the Linke turbidity coefficient. **Solar Energy**, v.81, p.896–903, 2007.
- OLIVEIRA, B. F. A.; IGNOTTI, E.; HACON, S. S. A systematic review of the physical and chemical characteristics of pollutants from biomass burning and combustion of fossil fuels and health effects in Brazil. **Cad. Saúde Pública**, Rio de Janeiro, v.27, p.1678-1698, 2011.

- PEDRÓS, R.; UTRILLAS, M. P.; MARTÍNEZ-LOZANO, J. A.; TENA, F. Values of broadband turbidity coefficients in a Mediterranean coastal site. **Solar Energy**, v.66, n.1, p.11-20, 1999.
- PORTILLO-QUINTERO, C.; SANCHEZ-AZOFEIFA, A.; ESPIRITO-SANTO, M. M. Monitoring deforestation with MODIS Active Fires in Neotropical dry forest: Na analysis of local-scale assessments in Mexico, Brazil and Bolivia. **Journal of Arid Environments**, v.97, p.150-159, 2013.
- POWER, H. C.; GOYAL, A. Comparison of aerosol and climate variability over Germany and South Africa. **Int. J. Climatol.** 23, p.921–941, 2003.
- RAHOMA, U. ALI.; HASSAN, A. H. Determination of atmospheric turbidity and its correlation with climatologically parameters. **American Journal of Environmental Science**, v.8, p.597-604, 2012.
- REBOITA, M. S.; GAN, M. A.; ROCHA, R. P.; AMBRIZZI, T. Regimes de precipitação na América do Sul: Uma revisão bibliográfica. **Revista Brasileira de Meteorologia**, v.25, p. 185-204, 2010.[in portuguese]
- REMER, L. A. ET AL. Global aerosol climatology from the MODIS satellite sensors. **Journal of Geophysical**, v. 113, p. 1-18, 2008.
- ROSENFELD, D. TRMM observed first direct evidence of smoke from forest fires inhibiting rainfall. **Geophysical Research Letters**, v.26, n.20, p.3105-3108, 1999.
- SALAZAR, G. A. Estimation of monthly values of atmospheric turbidity using measured values of global irradiation and estimated values from CSR and Yang Hybrid models. Study case: Argentina. **Atmospheric Environment**, v.45, p. 2465-2472, 2011.
- SATYAMURTY, P.; NOBRE, C.; SILVA DIAS, P. South America. In: D. Karoly and Vincent D., (eds.), Meteorology of the Southern Hemisphere, Meteorological Monographs 27 No. 49, pp. 119–139, **American Meteorological Society**, Boston , 1998.
- SCARAMBONI, C.; URBAN, R. C.; LIMA-SOUZA, M.; NOGUEIRA, R. F. P.; CARDOSO, A. A.; ALLEN, A. G.; CAMPOS, M. L. A. M. Total sugars in atmospheric aerosols: An alternative tracer for biomass burning. **Atmospheric Environment**, v.100, p.185-192, 2015.
- TERAMOTO, E. T.; ESCOBEDO, J. F. Análise da frequência anual das condições de céu em Botucatu, São Paulo. **Revista Brasileira Engenharia Agrícola Ambiental**. v.16, n. 9, p.985-992, 2012. [in portuguese]

TRABELSI, A.; MASMOUDI, M. An investigation of atmospheric turbidity over Kerkennah Island in Tunisia. **Atmospheric Research**, v.101, p.22–30, 2011.

USCKA-KOWALKOWSK, J. An analysis of the extinction of direct solar radiation on Mt. Kasprowy Wierch, Poland. **Atmospheric Research**, v.134, p.175–185, 2013.

VIDELA, F. C.; BARNABA, F.; ANGELINI, F.; GREMADES, P.; GOBBI, G. P. The relative role of Amazonian and non-Amazonian fires in building up the aerosol optical depth in South America: A five year study (2005-2009). **Atmospheric Research**, v.122, p.298-309, 2013.

WEN, C-C.; YEH, H-H. Analysis of atmospheric turbidity levels at Taichung Harbor near the Taiwan Strait. **Atmospheric Research**, v.94, p.168–177, 2009.

WMO – WORLD METEOROLOGICAL ORGANIZATION. **Guide to meteorological Instruments and Methods of Observation**. WMO-n°8, Seventh Edition, 2008, p.157-197, Geneva, Switzerland.

ZAKAY, A. S.; ABDELWAHAB, M. M.; MAKAR, P. A. Atmospheric turbidity over Egypt. **Atmospheric Environment**, v.38, p.1579–1591

8. CONCLUSÕES GERAIS

O trabalho apresentado avalia métodos de aprendizado de máquinas e modelos estatísticos na estimativa da irradiação direta na incidência normal. As técnicas utilizadas foram as Redes Neurais Artificiais e a Máquina de Vetores de Suporte. Diferentes variáveis de entrada foram testadas nas partições de tempo horárias e diárias.

As variáveis de entradas avaliadas mostraram que são capazes de modelar a H_b , contudo, os modelos com kt , n e H_{sc} tiveram melhores estimativas. Na estimativa de H_b com SVM, recomenda-se a utilização dos parâmetros: $C = 50$, $\gamma = 0,2$ e $\varepsilon = 0,005$ para a função de Kernel RBF. Nos modelos das redes neurais alterações nos parâmetros de entrada não resultaram em melhoria no desempenho dos modelos.

A partir dos resultados observados, as técnicas de aprendizado de máquinas foram capazes de modelar as diferentes combinações de dados de entrada na estimativa de H_b . Os modelos gerados a partir da SVM estimaram H_b com melhor exatidão e precisão que os modelos de RNA.

O desempenho dos modelos de SVM são superiores aos modelos estatísticos. Os modelos de RNA tiveram resultados semelhantes aos modelos estatísticos. Nesse caso, o uso do modelo é optativo, mas deixamos destacado a RNA como a primeira escolha com relação

aos modelos estatísticos. Por fim, a SVM é indicada como a primeira opção na estimativa de H_b .

A turbidez tem variabilidade ao longo do dia, aumenta de manhã e reduz à tarde, com máximos valores próximos ao meio dia solar. A variabilidade média horária diurna de TL é intensificada pela elevação da intensidade do fluxo radiante. Para a variabilidade diária anual, na primavera e no verão o vapor d'água na atmosfera e os aerossóis são os principais fatores responsáveis pela elevação da turbidez atmosférica. No verão o aumento da turbidez é causado, principalmente, pela elevação do vapor d'água na atmosfera.

O aumento das queimadas, a crescente urbanização e a expansão industrial local e regional são responsáveis pela tendência crescente da turbidez. Portanto, com o desmatamento de florestas nativas e a crescente urbanização, novos trabalhos devem ser realizados na perspectiva de analisar os novos padrões na atmosfera e na distribuição de DNI. Novas pesquisas tornam-se indispensáveis, pois estudos recentes mostram que queimadas acidentais e intencionais causados pelo homem têm aumento.

9. REFERÊNCIAS BIBLIOGRAFICAS

ALAM, S.; KAUSHIK, S. C.; GARG, S. N. Computation of beam solar radiation at normal incidence using artificial neural network. **Renewable Energy**, v.31, p.1483–1491, 2006.

ALLEN, R. G. Assessing integrity of weather data for reference evapotranspiration estimation. **Journal of Irrigation and Drainage Engineering**, v. 122, p.97-106, 1996.

ÅNGSTRÖM, A. K. Solar and terrestrial radiation. **Quarterly Journal of the Royal Meteorological Society**, v.50, p.121, 1924.

BATLLES, F. J.; RUBIO, M. A.; TOVAR, J.; OLMO, F. J.; ALADOS-ARBOLEDAS, L. Empirical modeling of hourly direct irradiance by means of hourly global irradiance. **Energy**, v. 25, p. 675-688, 2000.

BEHAR, O.; KHELLAF, A.; MOHAMMEDI, K. Comparison of solar radiation models and their validation under Algerian climate – The case of direct irradiance. **Energy Conversion and Management**, v.98, p.236–251, 2015.

BIRD, R. E.; HULSTROM, R. L. **Direct insolation models**. Golden, CO: Solar Energy Research Institute (SERI/NREL), 1980. (SERI/TR-335-344).

GUEYMARD, C. A. Direct solar transmittance and irradiance predictions with broadband models. Part II: validation with high quality measurements. **Solar Energy**, v. 74, p. 381–395, 2003.

GUEYMARD, C. A. **Progress in direct irradiance modeling and validation**. In: SOLAR 2010 CONF., 2010, Phoenix, AZ., American Solar Energy Soc, 2010. Disponível em: <http://www.solarconsultingservices.com/Gueymard-DNI_models_validation-ASES10.pdf>. Acesso em: 05 Março de 2015.

GUEYMARD, C. A. Temporal variability in direct and global irradiance at various time scales as affected by aerosols. **Solar Energy** (2012), doi:10.1016/j.solener.2012.01.013.

HALTHORE, R. N.; SCHWARTZ, S. E.; MICHALSKY, J. J.; ANDERSON, G. P.; FERRARE R. A.; HOLBEN B. N.; TEN BRINK H. M. Comparison of Model Estimated and Measured Direct-Normal Solar Irradiance. **Journal Geophysical Research**, v. 102, p.29991-30002, 1997.

INEICHEN, P. A broadband simplified version of the Solis clear sky model. **Solar Energy**, v.82, p.758-762, 2008.

IQBAL, M. **An introduction to solar radiation**. New York: Academic Press. 1983, 390p.

JIAN, Y. Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. **Energy**, v. 34, p.1276-1283, 2009.

KAUSHINA, N. D. TOMAR, R. K.; KAUSHIK, S. C. Artificial Neural Network model based on interrelationship of direct, diffuse and global solar radiations. **Solar Energy**, v. 103, p.327-342, 2014.

LIOU, K. N. **An Introduction to atmospheric radiation**. 2. Ed. San Diego: Academic Press, 2002, 583p.

LIU, B. Y. H.; JORDAN, R. C. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. **Solar Energy**, v.3, p.1-19, 1960.

LOUCHE, A.; NOTTON, G.; POGGI, P.; SIMONNOT, G. Correlations for direct normal and global horizontal irradiation on French Mediterranean site. **Solar Energy**, v.46, p.261-266, 1991.

LUNA, A. S.; PAREDES, M. L. L.; OLIVEIRA, G. C. G.; CORRÊA, S. M. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. **Atmospheric Environment**, v. 98, p.98-104, 2014.

MOHAMMADI, K.; SHAMSHIRBAND, S.; ANISI, M. H.; ALAM, K. A.; PETKOVIC, D. Support vector regression based prediction of global solar radiation on a horizontal surface. **Energy Conversion and Management**, v.91, p.433–441, 2015.

MONTEITH, J. L. Climate and the efficiency of crop production in Britain. **Philos. Trans. Royal London. London**, v.2081, p.277-294, 1977.

OLIVEIRA, A. P.; ESCOBEDO, J. F.; MACHADO, A. J. A new shadow-ring device for measuring diffuse solar radiation at the surface. **Journal of Atmospheric and Oceanic Technology**, v. 19, p. 698-708, 2001.

ORGILL, J. F.; HOLLANDS, K. G. T. Correlation equation for hourly diffuse radiation on a horizontal surface. **Solar Energy**, v.19: p.357–369, 1977.

PEREIRA, A. R.; ANGELOCCI, L. R.; SENTELHAS, P. C. **Agrometeorologia: fundamentos e aplicações**. Agropecuária, Guaíba, 2002, 478p.

RAGHAVENDRA, S.; DEKA, P. C. Support vector machine applications in the field of hydrology: A review. **Applied Soft Computing**, v.19, p.372–386, 2014.

REINDL, D. T.; BECKMAN W. A.; DUFIE, J. A. Diffuse fraction correlations. **Solar Energy**, v.45, p.1–7, 1990.

TOMAR, R. K.; KAUSHIKA, N. D.; KAUSHIK, S. C. Artificial neural network based computational model for the prediction of direct solar radiation in Indian zone. **Journal of Renewable and Sustainable Energy** 4, 063146 (2012); doi: 10.1063/1.4772677.

VAPNIK, V. N. **The nature of statistical learning theory**. New York: Springer, 1995.

YADAV, A. K.; CHANDEL, S. S. Solar radiation prediction using Artificial Neural Network techniques: A review. **Renewable and Sustainable Energy Reviews**, v.33, p. 772–781, 2014.

YANG, K.; HUANG, G.W.; TAMAI, N. A hybrid model for estimating global solar radiation. **Solar Energy**, v. 70, p. 13–22. 2001.