

03

Caracterização de Modelos para a Predição de Poluentes Atmosféricos

Jorge L. B. Ribes¹

Maicon Nardino²

Elisia R. Corrêa³

Régis S. Quadros⁴

Willian S. Barros⁵

Daniela Buske⁶

Programa de Pós-Graduação em Modelagem Matemática (PPGMMat - UFPel)

Pelotas – RS – Brasil

Resumo:

O objetivo do trabalho baseia-se na análise estatística dos resultados obtidos no desenvolvimento analítico das equações de advecção-difusão utilizando-se a combinação da Transformada de Laplace e da técnica GILTT. Os resultados obtidos por quatro modelos, utilizando os dados do experimento de Copenhagen, foram analisados. As análises estatísticas realizadas mostram que o modelo proposto é eficaz para a predição da concentração de poluentes utilizando o conjunto de dados do experimento de Copenhagen.

Abstract:

The objective of this study is based on statistical analysis of results obtained in the analytical development of the advection-diffusion equations using the combination of the Laplace transform and GILTT technique. The results obtained by four different models, using the Copenhagen experimental data, were analyzed. The performed statistical analyzes show that the proposed model is able to predict the concentration of pollutants using the Copenhagen experiment data set.

1. jlbribes@gmail.com

2. nardinomn@gmail.com,

3. elisiarc@hotmail.com

4. regisquadros@gmail.com

5. willianbarros@gmail.com

6. danielabuske@gmail.com

1 Introdução

Nossa atenção e percepção de que o planeta Terra enfrenta uma crise ambiental sem precedentes, que afeta a sobrevivência da humanidade, é muito recente e entrou no debate internacional mais efetivamente nas últimas décadas do século passado. Uma evolução sustentável implica numa gestão de responsabilidade socioambiental.

Em relação ao controle da poluição atmosférica em escala local ou regional geralmente se utiliza as redes de monitoramento da qualidade do ar. Esta rede constitui um instrumento útil para a segurança da saúde humana e do ambiente, e permite analisar o benefício de ações de saneamento e predispor intervenções específicas no caso de acontecer superação dos níveis do limiar estabelecido pela legislação (MOREIRA; TIRABASSI, 2004).

Na prática a gestão da qualidade do ar pode ser muito comprometida, pois o número de pontos de medida do monitoramento é limitado, além disso, a sua disposição espacial, pode não ter sido estudada detalhadamente e corre-se o risco de terem sido posicionadas em locais pouco representativos. Diante destes fatos, os modelos matemáticos que simulam o transporte e a difusão dos poluentes na atmosfera constituem uma opção importante para auxiliar as medidas de concentrações e prever a sua evolução. Assim, a utilização da modelagem matemática dos fenômenos de dispersão de poluentes produz um ganho significativo na qualidade na gestão da poluição atmosférica.

Diante disso, vários pesquisadores foram estimulados no desenvolvimento de modelos destes fenômenos de dispersão de poluentes. O grande desafio tem sido desenvolver e comparar os modelos que produzam resultados eficazes na estimação do nível de concentração de poluentes na atmosfera provenientes das inúmeras fontes de emissão, tanto antropogênicas como naturais.

Este trabalho baseia-se na comparação de modelos, por métodos estatísticos e análise estatística, dos resultados obtidos no desenvolvimento analítico das equações de advecção-difusão utilizando-se a combinação da Transformada de Laplace e da técnica GILTT (*Generalized Integral Laplace Transform Technique*). Esta técnica compreende os seguintes passos para a resolução de um problema bidimensional estacionário, por exemplo: solução de um problema associado de

Sturm-Liouville, expansão da concentração de poluentes em uma série em termos das autofunções, substituição desta expansão na equação de difusão-advecção e finalmente tomar momentos. Esse procedimento nos leva a um conjunto de equações diferenciais ordinárias chamadas de problema transformado. O problema transformado é resolvido pela técnica da Transformada de Laplace e diagonalização (WORTMANN et al., 2005). A técnica GILTT é um método totalmente analítico, no sentido de que nenhuma aproximação é feita ao longo da sua derivação exceto pelo erro de truncamento.

A modelagem do problema se dá na Camada Limite Planetária (CLP) que é definida como a parte da troposfera que é diretamente influenciada pela superfície terrestre e responde aos forçantes superficiais com uma escala de tempo inferior a uma hora, efetuando-se uma análise estatística dos resultados. Aplicaremos alguns conceitos estatísticos de acordo com os resultados encontrados pré-conhecidos na literatura e encontrados em (MOREIRA ET AL., 2009; BUSKE et al., 2012).

Os resultados foram gerados pela solução analítica da equação de advecção-difusão do tipo: bidimensional estacionário, bidimensional transiente, tridimensional considerando uma Gaussiana na direção transversal e, tridimensional totalmente analítico, utilizando os dados do experimento de Copenhagen (GRYNING; LICK, 1984; GRYNING et al., 1987).

A partir dos dados obtidos, além da análise pela técnica GILTT, serão realizadas análises dos estimadores erro quadrático médio normalizado (NMSE), coeficiente de correlação (COR), desvio fracional padrão (*FB*), fração de inclinação (*FS*), além de média, variância, desvio padrão, com a avaliação dos modelos com relação aos estimadores estatísticos analisados. Foi realizado ainda a análise estatística com base na análise da variância dos resíduos entre os valores observados e preditos pelos modelos.

A estatística descritiva compreende o conjunto dos métodos estatísticos para resumir e descrever as características proeminentes de conjuntos de dados referentes a observações de fenômenos particulares de uma classe de fenômenos. Nesse sentido, ela inclui a condensação de dados em forma de tabelas, sua representação gráfica e a determinação de indicadores numéricos de centro e de variabilidade (DA SILVA, 2004; PIMENTEL-GOMES, 2009).

A propriedade da estatística da entropia é interpretada como uma medida do grau de heterogeneidade do conjunto de dados observados e preditos de um sistema. A análise estatística através do uso da entropia permite uma interpretação da desordem dos elementos presentes nos modelos matemáticos, quanto mais próximo de zero (0) melhor o modelo prediz os valores observados (DEMÉTRIUS, 1978).

O presente artigo é uma versão estendida do artigo publicado por (NARDINO et al., 2016) no WCAMA de 2016. Neste presente trabalho foram incorporadas novas análises estatísticas dos dados, visando o aumento da acurácia dos resultados apresentados conforme a predição dos modelos GILTT. O artigo traz resultados referentes a estatística da entropia, como medida para verificação da heterogeneidade dos modelos GILTT, gráficos *box plot* para a discussão das estatísticas descritivas e análise de regressão, para verificação do ajuste entre os dados preditos e observados.

O presente artigo teve por objetivo de aplicar as medidas descritivas e efetuar o cálculo da entropia para verificar o grau de distorção entre os dados observados e preditos por quatro modelos matemáticos.

2 Modelagem

2.1 Modelagem Matemática

A equação de advecção-difusão da poluição do ar é uma representação da conservação de massa na atmosfera. Assumindo um fechamento Fickiano da turbulência, a equação de advecção-difusão pode ser escrita como (Buske, 2008):

$$\frac{\partial c}{\partial t} + \bar{u} \frac{\partial c}{\partial x} + \bar{v} \frac{\partial c}{\partial y} + \bar{w} \frac{\partial c}{\partial z} = \frac{\partial}{\partial x} \left(K_x \frac{\partial c}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_y \frac{\partial c}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_z \frac{\partial c}{\partial z} \right) + S \quad (1)$$

onde, c representa a concentração média do poluente em g/m^3 , \bar{u} , \bar{v} , \bar{w} representam as componentes do vento médio em m/s nas direções x , y e z , respectivamente, K_x , K_y e K_z são as componentes cartesianas do coeficiente de difusão nas direções x , y e z , respectivamente e S é o termo fonte. Na primeira ordem de fechamento, todas as informações sobre a turbulência do processo estão contidas nos coeficientes de difusividade turbulenta.

A equação (1) está sujeita as condições de contorno de fluxo nulo nos contornos:

$$K_x \frac{\partial c}{\partial x} = 0 \text{ em } x = 0 \text{ e } x = L_x \quad (2)$$

$$K_y \frac{\partial c}{\partial y} = 0 \text{ em } y = 0 \text{ e } x = L_y \quad (3)$$

$$K_x \frac{\partial c}{\partial x} = 0 \text{ em } z = 0 \text{ e } x = L_z \quad (4)$$

É considerado que a atmosfera está limpa no tempo inicial, ou seja, $c = 0$. A condição de fonte é representada por, onde E é a taxa de emissão em g/s , H a altura da CLP em metros, z_0 é a altura da fonte em metros, e $\delta(x - x_0)$ são as distâncias longe da fonte e δ representa a função delta de Dirac.

A solução do problema é obtida aplicando-se a técnica da transformada integral na variável y , transformando o problema 3D em um problema 2D resolvido em (Buske et al., 2012). Aplicando a transformada de Laplace na variável temporal, obtemos um problema estacionário, cuja solução é encontrada pelo método GILTT (Moreira et al., 2009).

2.2 Análises dos resultados dos modelos por métodos estatísticos

As análises estatísticas foram realizadas com intuito de verificar o comportamento do conjunto de dados observados e preditos pelos quatro modelos GILTT, neste sentido foram aplicadas medidas estatísticas descritivas.

Quanto às medidas de dispersão (variância, desvio padrão), estas indicam a distância que os valores se encontram da média e o quanto de variação o conjunto possui. O desvio padrão e a variância são dados, respectivamente por:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (5)$$

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (6)$$

Ainda, o coeficiente de variação, indica em termos percentuais, qual a variação dos dados, e é dado por:

$$CV = \frac{S}{\bar{x}} \times 100 \quad (7)$$

Erro quadrático médio normalizado (NMSE), é a média quadrática da diferença entre os valores observados (C_o) e preditos (C_p) que quando normalizado apresenta valores de 0 a 1. Tendo os menores erros os valores próximos a zero. Dado por:

$$\frac{(C_o - C_p)^2}{C_o C_p} \quad (8)$$

Coefficiente de Correlação (COR), revela a medida da intensidade da relação linear entre duas variáveis, o valor mais próximo a 1 na relação de C_o e C_p maior é o grau de ajuste linear entre os conjuntos. O coeficiente é estimado em razão da covariância entre os valores observados e preditos, e dividido pela multiplicação dos respectivos desvios.

$$\frac{(\text{COV}_{C_o p})}{\sqrt{\sigma_o^2 \sigma_p^2}} \quad (9)$$

Fator de dois (FA2), valores apresentam uma variação dentro do intervalo: $0.5 \leq \leq 2$. Quanto mais próximo de 1, mais próximos estão os dados. É dado por:

$$\frac{\overline{C_p}}{\overline{C_o}} \quad (10)$$

Desvio fracional padrão (FS), informa a tendência do modelo de predição de subestimar e sobreestimar os valores preditos. sendo, que valor ótimo é zero.

$$\frac{2(\sigma_o - \sigma_p)}{(\sigma_o + \sigma_p)} \quad (11)$$

Fração de inclinação (FB), que indica a tendência do modelo de superestimar ou subestimar as concentrações observadas. O valor ótimo é zero. É dado por:

$$\frac{2(\overline{C_o} - \overline{C_p})}{(\overline{C_o} + \overline{C_p})} \quad (12)$$

O cálculo da entropia foi realizado utilizando-se a variância dos resíduos entre os dados observados e preditos, pressupondo que os resíduos assumem distribuição normal. A entropia foi estimada pela equação 13:

$$\frac{1}{2} \ln(2\sigma^2 \pi e) \quad (13)$$

As análises estatísticas foram realizadas no programa computacional R (2015).

3 Resultados e Discussões

Os dados observados e preditos pelos modelos GILTT estão dispostos na Tabela 1. Foram utilizados os dados de concentração integrados lateralmente para a equação bidimensional estacionária (aqui denominado GILTT-1), dados com uma maior resolução temporal para a equação bidimensional transiente (aqui denominado GILTT-2), dados tridimensionais para a equação tridimensional estacionária que aqui é resolvida assumindo-se que em y tem-se uma distribuição Gaussiana da pluma como em (MOREIRA et al., 2009) (aqui denominado GILTT-3) e dados tridimensionais do modelo tridimensional analítico (aqui denominado GILTT-4). Podemos observar da Tabela 1 que o modelo GILTT descreve satisfatoriamente os dados observados no experimento de Copenhagen.

Na análise do modelo GILTT-1 para o conjunto de dados observados e preditos as medianas foram de 4.65 e 3.25 em $.10^{-4} \text{sm}^{-2}$. Com relação aos quartis, no primeiro conjunto de dados, verifica-se que o primeiro quartil tem valores 3.27 ($.10^{-4} \text{sm}^{-2}$) para os valores observados e 2.31 ($.10^{-4} \text{sm}^{-2}$) para os valores preditos, ou seja, 25% dos valores estão abaixo desses valores e que 75% deles estão acima, enquanto o terceiro quartil tem valores 6.84 e 5.84 ($.10^{-4} \text{sm}^{-2}$) para os valores observados e preditos, respectivamente.

Para o modelo GILTT-2 os valores das medianas para os dados observados e preditos foram 4.10 e 3.76 ($.10^{-4} \text{sm}^{-2}$). Tem-se no segundo conjunto de dados os valores 2.56 e 2.31 ($.10^{-4} \text{sm}^{-2}$) para o primeiro quartil para os dados observados e preditos, respectivamente e os valores 5.53 e 6.22 ($.10^{-4} \text{sm}^{-2}$) respectivamente, para os dados observados e preditos, para o terceiro quartil.

Com relação ao terceiro e quarto conjunto de dados no modelo GILTT-3 e GILTT-4 para os valores observados e preditos os valores das medianas foram 4.53 e 2.83 ($.10^{-4} \text{sm}^{-2}$). Para os quartis os valores foram de 2.38 e 2.14 ($.10^{-4} \text{sm}^{-2}$) para o primeiro quartil e 5.43 e 7.95 ($.10^{-4} \text{sm}^{-2}$) para o terceiro quartil.

Tabela 1 - Concentrações superficiais integradas lateralmente normalizadas pela taxa de emissão (.10-4sm-2), onde Cp são os dados preditos pelos modelos e Co representa as concentrações observadas experimentalmente.

Dados Co1	GILTT-1 Cp1	Dados Co2	GILTT-2 Cp2	Dados Co3	GILTT-3 Cp3	Dados Co4	GILTT-3 Cp4
6,48	6,84	6,48	5,53	10,5	4,53	10,5	12,22
2,31	3,97	2,31	3,4	2,14	1,7	2,14	4,25
5,38	4,65	5,38	3,46	9,85	5,41	9,85	8,23
2,95	3,05	2,95	2,47	2,83	2,12	2,83	2,86
8,2	8,14	8,2	7,47	16,33	8,83	16,33	16,49
6,22	5,19	6,22	4,78	7,95	3,68	7,95	5,66
4,3	3,98	4,3	3,77	3,76	2,26	3,76	3,29
11,66	9,24	11,66	9,02	15,71	14,3	15,71	10,44
6,72	8,58	6,72	8,98	12,11	17,69	12,11	16,24
5,84	6,71	5,84	6,31	7,24	8,55	7,24	7,09
4,97	5,39	4,97	5,04	4,75	5,43	4,75	3,96
3,96	3,51	-	-	7,44	5,75	7,44	6,80
2,22	2,5	-	-	3,37	2,38	3,37	2,39
1,83	1,98	-	-	1,74	1,5	1,74	1,49
6,7	4,67	6,7	3,53	9,48	4,25	9,48	8,66
3,25	2,76	3,25	2,37	2,62	1,53	2,62	2,67
2,23	2,23	2,23	1,95	1,15	1,05	1,15	1,70
4,16	4,83	4,16	4,45	9,76	5,99	9,76	10,04
2,02	3,27	2,02	3,17	2,64	2,75	2,64	4,04
1,52	2,63	1,52	2,56	0,98	1,77	0,98	2,23
4,58	4,45	4,58	3,18	8,52	5,02	8,52	7,83
3,11	2,91	3,11	2,29	2,66	1,95	2,66	2,72
2,59	2,19	2,59	1,8	1,98	1,15	1,98	1,47

Os valores (Tabela 2) para NMSE são aproximadamente zero, COR e FA2 tem valores próximos de 1. FB e FS são próximos de zero, indicando uma boa eficiência dos modelos GILTT, na predição dos valores (HANNA, 1989; PIMENTEL-GOMES, 2009).

Percebe-se também, que os dados observados e preditos possuem valores bastante semelhantes, o que nos remete a ideia da correlação, que demonstra a relação entre os dois conjuntos de dados (observados e preditos). No presente trabalho, remete ao quão próximo estão os resultados do nosso modelo GILTT, com os

valores reais observados, as magnitudes das correlações foram consideravelmente elevadas (COR = 0.909) no primeiro conjunto de dados, (COR = 0.848) no segundo, (COR=0,802) terceiro conjunto e (COR = 0,926) no quarto conjunto de dados, que estão muito próximo de 1, garantindo que os dados preditos estão altamente correlacionados com os valores observados. O modelo GILTT de predição pode estimar com bastante precisão o que será observado em coletas futuras, principalmente quando é utilizado o modelo GILTT-4. Os valores para NMSE são baixos, COR e FA2 tem valores próximos de 1 e FB e FS são próximos de zero,

garantindo uma boa performance do modelo utilizado.

Tabela 2 - Avaliação estatística utilizando o experimento de Copenhagen para os quatro modelos GILTT -1, -2, -3 e -4.

	NMSE*	COR	FA2	FB	FS
GILTT -1	0,050	0,909	1,000	-0,005	0,142
GILTT -2	0,090	0,848	1,000	0,107	0,127
GILTT -3	0,330	0,802	0,870	0,281	0,088
GILTT -4	0,070	0,926	0,957	0,019	0,026

*NMSE: erro quadrático médio, COR: coeficiente de correlação, FA2: fator de 2, FB: fator de inclinação e FS: desvio padrão fracional.

Na figura 1 são demonstrados os resultados da análise estatística dos dados referentes aos quatro modelos GILTT dos experimentos de Copenhagen, os quais os dados foram observados, e preditos pelos modelos. Com a análise estatística do conjunto de dados busca-se descrever sobre o grau de posição e variação entre os valores preditos e os dados que de fato foram observados.

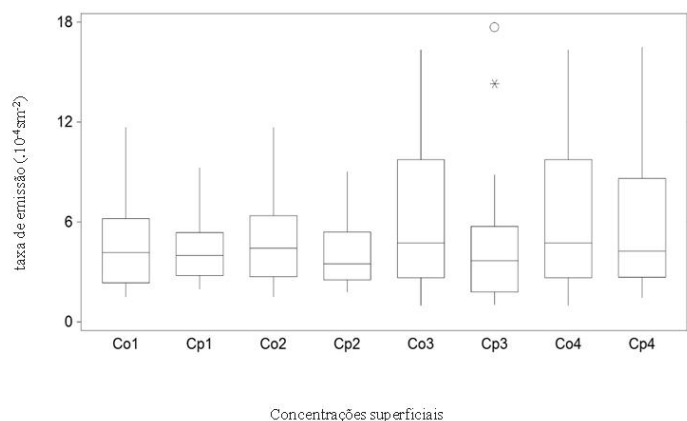


Figura 1. Gráfico box plot referente aos dados de Copenhagen observados e preditos pelos quatro modelos GILTT.

A entropia em estatística se refere a incerteza presente num conjunto de dados. Valores elevados de entropia remetem a presença de maior variabilidade no conjunto de dados, dificultando a obtenção de um modelo o qual melhor se ajusta ao conjunto de dados (Silva, 2014; DEMETRIUS, 1978). As estatísticas de entropia estimadas foram maiores para os modelos GILTT-3 (2,4493) e GILTT-4 (1,9789), podendo destacar que ocorre maior distorção entre os valores reais e preditos destes modelos (Figura 2), quando comparados com as estatísticas de entropia dos modelos GILTT-1 (1,4403) e GILTT-2 (1,6977) em que esta distorção é reduzida. Porém cabe salientar que para fins de compa-

ração de modelos, estas devem ser realizadas entre os modelos bidimensional e tridimensional, logo a comparação entre os dois modelos bidimensionais indica que o modelo GILTT-1 revelou menor variação que modelo GILTT-2. Na análise comparativa dos modelos tridimensionais o modelo GILTT-4 revelou estimativas de entropia inferiores ao modelo GILTT-3, assim pode-se inferir que a utilização dos modelos GILTT-2 e GILTT-4 são eficientes na predição de poluentes atmosféricos.

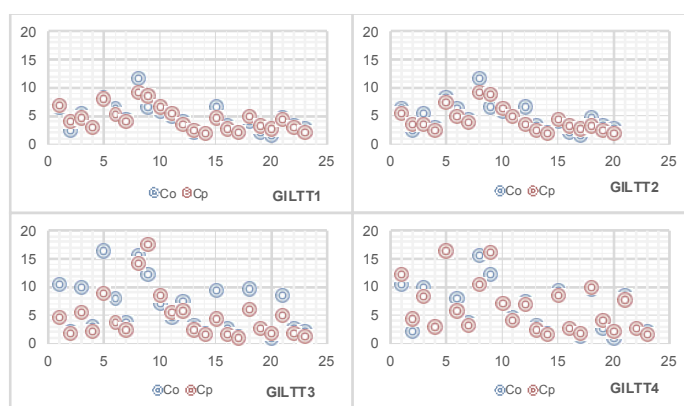


Figura 2 - Dispersão dos dados observados (Co) e preditos (Cp) nos quatro modelos GILTT.

Com relação à Figura 2, que demonstra a dispersão dos dados observados e preditos pelos modelos GILTT, pode-se inferir que para os modelos bidimensionais os ajustes foram semelhantes quanto a dispersão dos dados observados e preditos nos modelos GILTT-1 e GILTT-2. Os modelos GILTT-3 e GILTT-4 relacionados a estrutura de modelos tridimensionais também apresentaram dispersão semelhante entre os valores observados e preditos pelos modelos, mas pode-se inferir que o modelo GILTT-4 demonstrou pouca distorção na predição de valores de componentes da taxa de emissão do poluente hexafluoreto de enxofre em (10^{-4}sm^{-2}).

4 Conclusões

Nos quatro conjuntos de dados percebe-se que os valores de NMSE, COR, FA2, FS e FB levam-nos a acreditar que o modelo GILTT tem uma boa *performance* e mostrou-se confiável na predição da dispersão de poluentes atmosféricos, o que destaca a grande importância e eficácia no auxílio à prevenção de futuros problemas ligados à dispersão de poluentes na atmosfera. Percebe-se também que os valores observados e preditos apresentaram grande variação para as diferentes

coletas, indicando assim que tais coletas devem ser feitas em locais e horários diferenciados para uma melhor análise geral da dispersão de poluentes atmosféricos.

A estatística da entropia revela que entre os modelos bidimensionais o modelo GILTT-1 apresenta menor variância residual comparativamente ao modelo GILTT-2.

A estatística da entropia revela que entre os modelos tridimensionais o GILTT-4 tem menor variância residual, em comparação ao modelo GILTT-3.

Referências

- BUSKE, D. Solução *GILTT Bidimensional em Geometria Cartesiana*: Simulação da dispersão de poluentes na atmosfera, Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, 2008.
- BUSKE, D. et al. Air pollution steady-state advection-diffusion equation: the general three-dimensional solution. *Journal of Envir. Protec*, v.4, p.1-10, 2012.
- DA SILVA, J. G. C. *Estatística Básica*. Versão preliminar. Universidade Federal de Pelotas-UFPeL. Pelotas-RS, 2004.
- DEMETRIUS, L. Adaptive value, entropy and survivorship curve. *Nature*, v.275, p. 213–214, 1978.
- GRYNING, S. E.; LYCK, E. Atmospheric dispersion from elevated source in an urban area: comparison between tracer experiments and model calculations. *Journal of Climate applied Meteorology*, v.23, p. 651-654, 1984.
- GRYNING, S. E. et al. Applied dispersion modelling based on meteorological scaling parameters. *Atmospheric Environment*, v.21, n.79-89, 1987.
- HANNA, S. R. Condense limit for air quality models as estimated by bootstrap and jackknife resampling methods. *Atmospheric Environment*, v.23, p. 1385-1395, 1989.
- MOREIRA, D. M.; TIRABASSI, T. Modelo matemático de dispersão de poluentes na atmosfera: um instrumento técnico para a gestão ambiental. *Ambiente & Sociedade*, Campinas, v. 7, n. 2, p. 159-171, 2004.
- MOREIRA, D. M. et al. The state-of-art of the GILTT method to simulate pollutant dispersion in the atmosphere. *Atmospheric Research*, v.92, p. 1-17, 2009.
- NARDINO, M., Ribes, J., Correa, E. R., Quadros, R. S., Barros, W. S., Buske, D. (2016) Caracterização de Modelos para a Predição de Poluentes Atmosféricos. XXXVI Congresso da Sociedade Brasileira de Computação. In: WCAMA - WORKSHOP DE COMPUTAÇÃO APLICADA À GESTÃO DO MEIO AMBIENTE E RECURSOS NATURAIS. 7. 2016. Porto Alegre. *Anais...* Porto Alegre, 2016. p. 1858-1866. ISBN 978-85-7669-322-2.
- PIMENTEL-GOMES, F. *Curso de estatística experimental*. Piracicaba, v. 15, p. 451. 2009.
- R VERSION 3.2.2. *Fire Safety* Copyright (C) The R Foundation for Statistical Computing, 2015.
- WORTMANN, S. et al. A new analytical approach to simulate the pollutant dispersion in the PBL. *Atmospheric Environment*. v.39, n.12, p. 2187-2194, 2005.