

Caracterização de Modelos para a Predição de Poluentes Atmosféricos

**Maicon Nardino, Jorge Ribes, Elisia R. Correa, Régis S. Quadros,
Willian S. Barros, Daniela Buske**

Programa de Pós-Graduação em Modelagem Matemática (PPGMMat - UFPel)
Pelotas – RS – Brasil

{nardinomn, jlbribes, regisquadros, willianbarros,
danielabuske}@gmail.com, {elisiarc}@hotmail.com

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.*

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.*

1. Introdução

A qualidade do ar das cidades é um fator fundamental para a qualidade de vida da população e do meio ambiente. Devido a isso, diferentes grupos de pesquisa trabalham na modelagem de fenômenos atmosféricos e em sua simulação computacional. O grande desafio tem sido desenvolver ferramentas numéricas eficazes para estimar o nível de concentração de poluentes na atmosfera provenientes das inúmeras fontes de emissão, tanto antropogênicas como naturais. A partir da estimativa do nível de concentração próximo da fonte, pode-se realizar uma avaliação do impacto ambiental gerado e agir na tentativa de uma solução o mais conveniente possível para o problema.

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].

Este 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 (*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] e [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 and Lick, 1984] [Gryning et al., 1987].

Quanto às medidas de dispersão, o valor central do conjunto de dados, sendo uma estatística de posição, o desvio padrão e o coeficiente de variação. O desvio padrão (em dados brutos) e a variância são dados, respectivamente por:

$$S = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \text{ e } S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1},$$

sendo, que estes indicam, respectivamente, a distância que os valores se encontram da média e o quanto de variação o conjunto possui. Ainda, o coeficiente de variação, indica em termos percentuais, qual a variação dos dados, e é dado por:

$$CV = \frac{\sum_{i=1}^n S}{\bar{x}} \cdot 100$$

A partir dos dados obtidos, além da análise pela técnica GILTT, serão realizadas análises de *NMSE*, *COR*, *FB*, *FS*, além de média, variância, desvio padrão, correlação e distribuição de frequências, com a avaliação dos modelos com relação aos estimadores estatísticos analisados.

2. Modelagem Matemática do Problema

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, $c(x,y,z,t)$ 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:

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

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

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

É considerado que a atmosfera está limpa no tempo inicial, ou seja, $c(x,y,z,0)=0$. A condição de fonte é representada por $\bar{c}(0,y,z,t)=Q\delta(y-y_0)\delta(z-H_s)$, onde Q é a taxa de emissão em g/s, z_i a altura da CLP em metros, H_s é a altura da fonte em metros, L_x e L_y 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].

3. Resultados

Os resultados preditos neste trabalho foram obtidos pelo método da GILTT à partir de dados coletados em Copenhagen e foi feita uma análise estatística quanto as medidas de posição e dispersão, tais como, média, mediana, desvio padrão, simetria, curtose, variância e coeficiente de variação e uma análise de correlação, além de uma análise de distribuição de frequência. Os detalhes do experimento de Copenhagen estão descritos em [Buske, 2008], [Gryning and Lick, 1984], [Gryning et al., 1987].

Na Tabela 1 podemos ver as concentrações observadas no experimento e preditas pelo modelo GILTT para os quatro casos estudados. 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 [7] (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.

Tabela 1. Concentrações superficiais integradas lateralmente normalizadas pela taxa de emissão ($\cdot 10^{-4} \text{sm}^{-2}$), onde C_p são os dados preditos pelos modelos e C_o representa as concentrações observadas experimentalmente.

Dados C_{o1}	GILTT-1 C_{p1}	Dados C_{o2}	GILTT-2 C_{p2}	Dados C_{o3}	GILTT-3 C_{p3}	Dados C_{o4}	GILTT-3 C_{p4}
6,48	6,84	6,48	5,53	10,5	4,53	10,50	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	6,7	3,53	7,44	5,75	7,44	6,80
2,22	2,5	3,25	2,37	3,37	2,38	3,37	2,39
1,83	1,98	2,23	1,95	1,74	1,5	1,74	1,49
6,7	4,67	4,16	4,45	9,48	4,25	9,48	8,66
3,25	2,76	2,02	3,17	2,62	1,53	2,62	2,67
2,23	2,23	1,52	2,56	1,15	1,05	1,15	1,70
4,16	4,83	4,58	3,18	9,76	5,99	9,76	10,04
2,02	3,27	3,11	2,29	2,64	2,75	2,64	4,04
1,52	2,63	2,59	1,8	0,98	1,77	0,98	2,23
4,58	4,45			8,52	5,02	8,52	7,83
3,11	2,91			2,66	1,95	2,66	2,72
2,59	2,19			1,98	1,15	1,98	1,47

Na Tabela 2 apresentam-se os resultados dos índices estatísticos, descritos por [Hanna, 1989], [Pimentel-Gomes, 2009], e definidos como:

Erro quadrático médio normalizado (NMSE), dado por:

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

que informa sobre todos os desvios das concentrações do modelo e as concentrações observadas. É uma medida adimensional e quanto menor o valor, melhor o modelo.

Coefficiente de Correlação (COR):

$$\frac{\overline{(C_o - \overline{C_o})(C_p - \overline{C_p})}}{\sigma_o \sigma_p}$$

que descreve o grau de associação ou concordância entre as variáveis. O valor 1 garante uma boa performance.

Fator de dois (FA2), fração dos dados (% normalizados a 1), entre 0.5 e 2, ou seja, $0.5 \leq \frac{C_o}{C_p} \leq 2$. Quanto mais próximo de 1, mais próximos estão os dados.

Desvio fracional padrão (FS), é dado por:

$$\frac{\sigma_o - \sigma_p}{0.5(\sigma_o + \sigma_p)}$$

sendo, que valor ótimo é zero.

Fração de inclinação (FB), é dado por:

$$\frac{\overline{C_o} - \overline{C_p}}{0.5(\overline{C_o} + \overline{C_p})}$$

que indica a tendência do modelo de superestimar ou subestimar as concentrações observadas. O valor ótimo é zero.

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 tabela 3 apresentam-se as análises estatísticas descritivas realizadas, que são o objetivo principal deste artigo.

Tabela 3. Tabela de análises estatísticas dos dados de Copenhagen observados e preditos (Co₁, Cp₂) para experimento 1, 2, 3 e 4.

Parâmetros	Experimento 1		Experimento 2		Experimento 3		Experimento 4	
	Co ₁	Cp ₁	Co ₂	Cp ₂	Co ₃	Cp ₃	Co ₄	Cp ₄
Somatório	103,20	103,67	95,19	85,53	145,51	109,59	145,51	142,77
Média	4,48	4,50	4,75	4,27	6,32	4,76	6,33	6,21
Mediana	4,65	3,25	4,1	3,765	4,53	2,83	4,53	2,83
Var*	5,98	4,50	6,19	4,89	21,19	17,77	21,20	20,12
DP	2,39	2,07	2,42	2,14	4,50	4,12	4,50	4,39
Coef. Var (%)	42,51	67,48	48,67	64,31	84,56	81,45	84,56	70,68
Simetria	0,56	1,91	0,94	1,42	2,18	1,17	2,19	1,17
Curtose	-0,62	4,29	0,13	2,16	5,57	1,09	5,57	1,09
Quartis 1	3,27	2,31	2,56	2,31	2,38	2,14	2,38	2,14
Quartis 3	6,84	5,84	5,53	6,22	5,43	7,95	5,43	7,95
Correlação	0,9096		0,8016		0,8016		0,93	

*Var: variância, DP: desvio padrão e Coef. Var (%): coeficiente de variação.

4. Discussões

Analisando a Tabela 3, foram observadas algumas propriedades dos dados através resultados estatísticos. No primeiro conjunto de dados, no que diz respeito às medidas de tendência central, foram verificadas a média e a mediana. Calculando-se a média encontram-se $\bar{x} = 4.4870$ e $\bar{x} = 4.5074$ para os valores observados e preditos, respectivamente, identificados como Co_1 e Cp_1 para Co_2 e Cp_2 , da mesma forma, tem-se $\bar{x} = 4.7595$ e $\bar{x} = 4.2765$, $\bar{x} = 6.3265$ e $\bar{x} = 4.7648$, identificados como Co_3 e Cp_3 , e ainda $\bar{x} = 6.33$ e $\bar{x} = 6.21$, identificados como Co_4 e Cp_4 , estabelecendo-se assim o valor médio de cada conjunto de dados. Encontram-se também as medianas $Me = 4.65$ e $Me = 3.25$, $Me = 4.1$ e $Me = 3.765$, $Me = 4.53$ e $Me = 2.83$ e $Me = 4.53$ e $Me = 2.83$, respectivamente para $(Co_1$ e $Cp_1)$, $(Co_2$ e $Cp_2)$, $(Co_3$ e $Cp_3)$ e $(Co_4$ e $Cp_4)$, que determinam, em cada conjunto, o valor central.

Pelas medidas de dispersão (Tabela 3) observa-se que a variância para o experimento 1 foi de 5.9862 para os dados observados e 4.5087 para os preditos, no experimento 2, 6.1965 para os dados observados e 4.8942 para os preditos, no experimento 3, 21.1956 para os dados observados e 17.7735 para os preditos e no modelo 4, 21.20 para os dados observados e 20.12 para os preditos, o que é bastante alto nos quatro experimentos, mostrando que possuem realmente uma grande diferença entre os seus valores. Nota-se que a variância dos dados, observados e preditos nos dá a nítida percepção de que a variância dos primeiros é maior que a dos preditos em todas as observações. Quanto ao desvio padrão, tem-se para os dados observados e preditos, nos quatro experimentos, respectivamente, 2.3929 e 2.0767, 2.4263 e 2.1462, 4.5027 e 4.1232, 4.50 e 4.39, e na mesma ordem os valores percentuais aproximados respectivos ao coeficiente de variação, 42.51 e 67.48, 48.67 e 64.31, 84.56 e 81.45, 84.56 e 70.68, garantindo assim, que a técnica GILTT é eficiente, gerando bons resultados. 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), ou seja, 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 ($COR = 90,96\%$) no primeiro conjunto de dados e ($COR = 80,16\%$) no segundo e terceiro conjuntos e ($COR = 93\%$) no quarto conjunto de dados, que estão muito próximo de 100%, garantindo que os dados preditos estão perto dos observados, isto é, que o modelo GILTT de predição pode estimar com bastante precisão o que será observado em coletas futuras. Os valores para $NMSE$ são bastante pequenos, COR e $FA2$ tem valores próximos de 1 e FB e FS são próximos de zero, garantindo uma boa performance do modelo utilizado.



Figura 1. Análise da distribuição de frequência para os dados observados (A1 e preditos (A2) pelo modelo GILTT 1, GLITT 2 (B1 e B2), GLITT 3 (C1 e C2) e GLITT (D1 e D2).

Pode-se também observar os quartis, sendo o primeiro quartil (Q_1) o valor que divide o conjunto em duas partes, tais que um quarto ou vinte e cinco por cento (25%) dos valores sejam menores do que ele e três quartos ou setenta e cinco por cento (75%) dos restantes sejam maiores. O elemento que indica a ordem ou posição do primeiro quartil é determinado, para dados agrupados em classes, pela expressão $E_{Q_1} = n/4$, em que n é o número de valores do conjunto, ou número de observações. O terceiro quartil é o valor que divide o conjunto em duas partes tais que um quarto ou vinte e cinco por cento (25%) dos valores sejam maiores do que ele e três quartos ou setenta e cinco por cento

(75%) dos restantes sejam menores. O elemento que indica a ordem ou posição do primeiro quartil é determinado, para dados agrupados em classes, pela expressão $E_{Q_3} = 3n/4$, em que n é o número de valores do conjunto, ou número de observações.

Na Tabela 3, no primeiro conjunto de dados, verifica-se que o primeiro quartil tem valores 3.2700 para os valores observados e 2.3100 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 6.8400 e 5.8400 para os valores observados e preditos, respectivamente, indicando assim que 75% dos valores estão abaixo e que 25% dos valores estão acima desses números. Indicando tal observação, tem-se no segundo conjunto de dados os valores 2.5600 e 2.3100 para o primeiro quartil para os dados observados e preditos, respectivamente e os valores 5.5300 e 6.2200 respectivamente, para os dados observados e preditos, para o terceiro quartil. Analogamente, no terceiro conjunto de dados tem-se 2.3800 e 2.1400 para o primeiro quartil e 5.4300 e 7.9500 para o terceiro quartil, bem como no quarto conjunto de dados tem-se 2.3800 e 2.1400 para o primeiro quartil e 5.4300 e 7.9500 para o terceiro quartil, na ordem dos conjuntos de dados anteriores.

Com relação a análise de distribuição de frequência, como pode ser visto na Figura 1, onde os dados obtidos através das coletas são indicados por $A1$, $B1$, $C1$ e $D1$ e os dados obtidos pelas técnicas são indicados por $A2$, $B2$, $C2$ e $D2$, observa-se que houve a formação de quatro classes para os dados observados e também preditos. A classe com maior percentual de observações para modelo GILTT-1 reúne aproximadamente 45% das informações, tanto observadas como preditas, para o modelo GILTT-2 a classe com maior percentual de observações reúne 30% das informações observadas e preditas, para o modelo GILTT-3 a classe com maior percentual de observações reúne em torno 35% das informações observadas e preditas e para o modelo GILTT-3 a classe com maior percentual de observações reúne aproximadamente 25% das informações observadas e preditas.

5. Conclusões

Utilizou-se neste trabalho a solução analítica para a equação de advecção-difusão obtida pelo método GILTT para quatro problemas, com modelagem do problema na CLP e efetuou-se uma análise estatística dos resultados. A estatística descritiva foi utilizada para resumir e descrever as características dos conjuntos de dados, fundamentando-se na análise estatística dos resultados obtidos no desenvolvimento analítico das equações de advecção-difusão combinando o método da Transformada de Laplace e da técnica GILTT com os dados do experimento de Copenhagen.

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 leva-nos a entender que se de 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.

6. Referências

- Buske, D. (2008) Solução GILTT Bidimensional em Geometria Cartesiana: Simulação da dispersão de poluentes na atmosfera, Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul.
- Buske, D., Vilhena, M. T., Tirabassi, T. and Bodmann, B. (2012) “Air pollution steady-state advection-diffusion equation: the general three-dimensional solution”. *Journal of Envir. Protec*, 4:1-10.
- da Silva, J. G. C. (2004) “Estatística Básica”. Versão preliminar. Universidade Federal de Pelotas-UFPeL. Pelotas-RS.
- Gryning, S. E. and Lyck, E. (1984) “Atmospheric dispersion from elevated source in an urban area: comparison between tracer experiments and model calculations”. *Journal of Climate applied Meteorology*, 23: 651-654.
- Gryning, S. E., Holtslag, A.M.M., Irwin, J. and Sivertsen, B. (1987) “Applied dispersion modelling based on meteorological scaling parameters”. *Atmospheric Environment*, 21: 79-89.
- Hanna, S. R. (1989) “Condence limit for air quality models as estimated by bootstrap and jackknife resampling methods”. *Atmospheric Environment*, 23: 1385-1395.
- Moreira, D. M., Vilhena, M. T., Buske, D. and Tirabassi, T. (2009) “The state-of-art of the GILTT method to simulate pollutant dispersion in the atmosphere”. *Atmospheric Research* 92: 1-17.
- Pimentel-Gomes, F. (2009) “Curso de estatística experimental”. Piracicaba, v. 15: 451.
- Wortmann, S., Vilhena, M. T., Moreira, D. M. and Buske, D. (2005) “A new analytical approach to simulate the pollutant dispersion in the PBL”. *Atmospheric Environment*. 39 (12): 2187-2194.