

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Uma Análise Estatística de Dados Aplicado ao Estudo da Dispersão de Poluentes Atmosféricos

Jorge Ribes¹

Willian Barros²

Régis Sperotto de Quadros³

Daniela Buske⁴

Elisia Rodrigues Corrêa⁵

Maicon Nardino⁶

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

Resumo. 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. Os resultados obtidos pelos modelos bidimensional estacionário e transiente e tridimensional aproximado, 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.

Palavras-chave. Advecção-Difusão, GILTT, Dispersão de Poluentes, Modelo Analítico, Análise Estatística

1 Introdução

Nesse trabalho, utiliza-se a solução analítica para a equação de advecção-difusão obtida pelo método GILTT (*Generalized Integral Laplace Transform Technique*) para os problemas em equações do tipo: bidimensional estacionário, bidimensional transiente e tridimensional (considerando uma Gaussiana na direção transversal). 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 [7].

¹jlribes@gmail.com

²willianbarros@gmail.com

³regisquadros@gmail.com, Autor-apresentador

⁴danielabuske@gmail.com

⁵elisiarc@hotmail.com

⁶nardinomn@gmail.com

A técnica GILTT é um método totalmente analítico. Esta técnica compreende os seguintes passos: 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. Esse problema é resolvido pela técnica da Transformada de Laplace e diagonalização [9].

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 [8]. 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 [3]. 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, descrito acima. Serão apresentados resultados obtidos pelos modelos bidimensional estacionário e transiente e tridimensional aproximado, utilizando os dados do experimento de Copenhagen [4,5].

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, pode-se escrever a equação tridimensional transiente da seguinte forma:

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

onde $c(x, y, z, t)$ representa a concentração média de poluente (g/m^3), $\bar{u}, \bar{v}, \bar{w}$ representam as componentes do vento médio (m/s) nas direções x, y, z , respectivamente, K_x, K_y e K_z são as componentes cartesianas do coeficiente de difusão nas direções x, y, z , respectivamente, 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:

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

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

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

É considerado que a atmosfera está limpa no tempo inicial, ie, $c(x, y, z, 0) = 0$.

A condição de fonte é representada por $\bar{u}c(0, y, z, t) = Q\delta(y - y_o)\delta(z - H_s)$, onde Q é a taxa de emissão (g/s), z_i a altura da CLP (m), H_s é a altura da fonte (m), L_x e L_y são distâncias longe da fonte e δ representa a função delta de Dirac.

A solução do problema (1) é obtida aplicando-se a técnica da transformada integral na variável y , transformando o problema 3D em um problema 2D resolvido em [2]. Aplicando a transformada de Laplace na variável temporal, obtemos um problema estacionário, cuja solução é encontrada pelo método GILTT [7].

3 Resultados e discussões

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. Os detalhes do experimento de Copenhagen estão descritos em [1,4,5].

Na Tabela 1 podemos ver as concentrações observadas no experimento e preditas pelo modelo GILTT. 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), e 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).

Na Tabela 2 apresentam-se os resultados dos índices estatísticos, descritos por [6], e definidos como: Erro quadrático médio normalizado ($NMSE$), dado por $\frac{(C_o - C_p)^2}{C_o C_p}$ 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. Coeficiente de Correlação (COR), $\frac{(C_o - \bar{C}_o)(C_p - \bar{C}_p)}{\sigma_o \sigma_p}$ 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), que estão entre $0.5 \leq \frac{C_p}{C_o} \leq 2$. Quanto mais próximo de 1, mais próximos estão os dados. Desvio fracional padrão (FS), é dado por $\frac{\sigma_o - \sigma_p}{0.5(\sigma_o + \sigma_p)}$. O valor ótimo é zero. Fração de inclinação (FB), é dado por $\frac{\bar{C}_o - \bar{C}_p}{0.5(\bar{C}_o + \bar{C}_p)}$. Indica a tendência do modelo de superestimar ou subestimar as concentrações observadas. O valor ótimo é zero.

Podemos observar das Tabelas 1 e 2 que o modelo GILTT descreve satisfatoriamente os dados observados no experimento de Copenhagen.

A partir de agora serão feitas as análises estatísticas que são o objetivo principal deste artigo.

Analisando a Tabela 3, observam-se algumas propriedades dos dados através resultados estatísticos. Tem-se 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 C_{o1} e C_{p1} para C_{o2} e C_{p2} , da mesma forma, tem-se $\bar{x} = 4.7595$ e $\bar{x} = 4.2765$ e ainda $\bar{x} = 6.3265$ e $\bar{x} = 4.7648$, identificados como C_{o3} e C_{p3} , estabelecendo-se assim o valor médio de cada conjunto de dados. Encontram-se também as medianas $M_e = 4.65$ e $M_e = 3.25$, $M_e = 4.1$ e $M_e = 3.765$ e $M_e = 4.53$ e $M_e = 2.83$, respectiva-

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

GILTT-1		GILTT-2		GILTT-3	
Dados C_{o1}	Dados C_{p1}	Dados C_{o2}	Dados C_{p2}	Dados C_{o3}	Dados C_{p3}
6.48	6.84	6.48	5.53	10.50	4.53
2.31	3.97	2.31	3.40	2.14	1.70
5.38	4.65	5.38	3.46	9.85	5.41
2.95	3.05	2.95	2.47	2.83	2.12
8.20	8.14	8.20	7.47	16.33	8.83
6.22	5.19	6.22	4.78	7.95	3.68
4.30	3.98	4.30	3.77	3.76	2.26
11.66	9.24	11.66	9.02	15.71	14.30
6.72	8.58	6.72	8.98	12.11	17.69
5.84	6.71	5.84	6.31	7.24	8.55
4.97	5.39	4.97	5.04	4.75	5.43
3.96	3.51	6.70	3.53	7.44	5.75
2.22	2.50	3.25	2.37	3.37	2.38
1.83	1.98	2.23	1.95	1.74	1.50
6.70	4.67	4.16	4.45	9.48	4.25
3.25	2.76	2.02	3.17	2.62	1.53
2.23	2.23	1.52	2.56	1.15	1.05
4.16	4.83	4.58	3.18	9.76	5.99
2.02	3.27	3.11	2.29	2.64	2.75
1.52	2.63	2.59	1.80	0.98	1.77
4.58	4.45			8.52	5.02
3.11	2.91			2.66	1.95
2.59	2.19			1.98	1.15

Tabela 2: Avaliação estatística utilizando o experimento de Copenhagen

Modelo	NMSE	COR	FA2	FB	FS
GILTT-1	0.05	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

Tabela 3: Tabela das análises estatísticas.

Parâmetros	GILTT-1		GILTT-2		GILTT-3	
	C_{o_1}	C_{p_1}	C_{o_2}	C_{p_2}	C_{o_3}	C_{p_3}
Somatório	103.2000	103.6700	95.1900	85.5300	145.5100	109.5900
Média	4.4870	4.5074	4.7595	4.2765	6.3265	4.7648
Mediana	4.65	3.25	4.1	3.765	4.53	2.83
Variância	5.9862	4.5087	6.1965	4.8942	21.1956	17.7735
Desvio Padrão	2.3929	2.0767	2.4263	2.1462	4.5027	4.1232
Coef. de Var.(%)	42.5145	67.4847	48.6700	64.3100	84.5600	81.4540
Simetria	0.5616	1.9128	0.9412	1.4240	2.1897	1.1734
Curtose	-0.6280	4.2911	0.1362	2.1650	5.5724	1.0927
Quartil 1	3.2700	2.3100	2.5600	2.3100	2.3800	2.1400
Quartil 3	6.8400	5.8400	5.5300	6.2200	5.4300	7.9500
Correlação	0.9096		0.8016		0.8016	

mente para $(C_{o_1}$ e C_{p_1}), $(C_{o_2}$ e $C_{p_2})$, $(C_{o_3}$ e $C_{p_3})$ que determinam, em cada conjunto, o valor central. Quanto às medidas de dispersão, verificaram-se a variância, o desvio padrão e o coeficiente de variação, o desvio padrão (dado por $S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$, devido aos dados serem brutos), e a variância (dada por $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$), sendo que estes últimos indicam, respectivamente, a distância que os valores se encontram da média e o quanto de variação tem dentro desse conjunto. Ainda, o coeficiente de variação (dado por $CV = \frac{S}{\bar{x}} \cdot 100$), indica, em termos percentuais, qual a variação dos dados, que tanto nos dados observados quanto nos preditos é bastante alto nas três observações, mostra que possuem realmente uma grande diferença entre os seus valores. Nota-se que a variância dos dados, observados e preditos nos dão a nítida percepção de que a variância dos primeiros é maior que a dos preditos em todas as observações, garantindo assim, que a técnica GILTT é eficiente, gerando bons resultados. Percebe-se ainda, que os dados observados e preditos possuem valores bastante semelhantes, o que nos lembra a correlação que demonstra a relação entre os dois conjuntos de dados, 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 ($COR = 80, 16\%$) nos outros dois, 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.

Observando-se os gráficos da Figura 1 percebe-se para quantidade de cada valor eixo x varia uma certa quantidade no valor do eixo y , o que lembra-nos a regressão linear, que neste caso é positiva e mostra que toda vez que se modificar um valor observado o valor predito deverá acompanhar essa alteração. Isso acontece para todos, indicando assim a

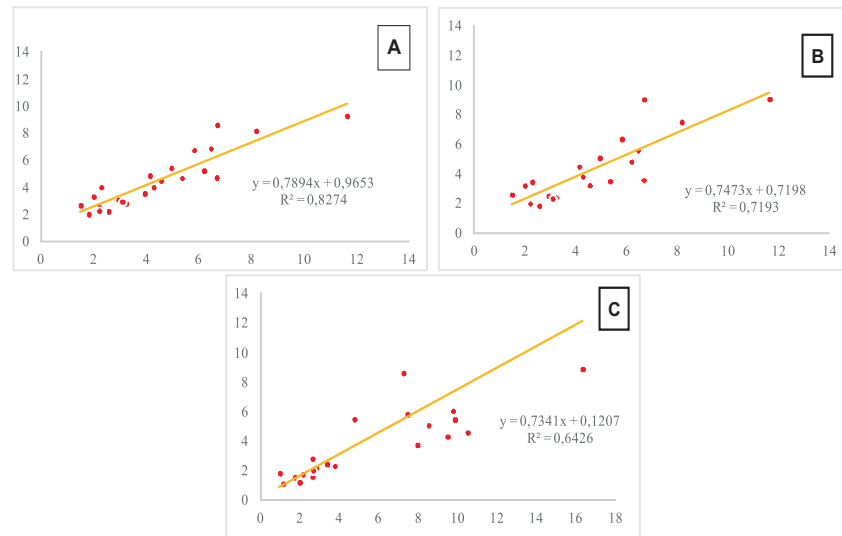


Figura 1: Análise de regressão entre os valores observados e os valores preditos em A (primeira coleta de dados), B (segunda coleta de dados) e C (terceira coleta de dados), referentes as coletas em Copenhagen.

estabilidade do modelo para esse experimento.

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 $E_{Q_1} = \frac{n}{4}$, 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 $E_{Q_3} = \frac{3n}{4}$, 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 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, na ordem dos conjuntos de dados anteriores.

4 Conclusões

Utilizou-se a solução analítica para a equação de advecção-difusão obtida pelo método GILTT para os 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 três 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 e também percebe-se que os valores observados e preditos apresentaram grande variação para as diferentes coletas.

Referências

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