

IMPLEMENTAÇÃO DO MÉTODO *KERNEL* DE DENSIDADE EM UM MODELO LAGRANGIANO DE TRAJETÓRIA DE PARTÍCULAS APLICADO A CORPOS D'ÁGUA

*Felipe Nascimento Gaze*¹; *Julio Tomás Aquije Chacaltana*² & *Maxsuel Marcos Rocha Pereira*³.

RESUMO --- Um algoritmo relacionado ao método do *kernel* de densidade para cálculo da concentração de um dado poluente é implementado em um Modelo Lagrangiano de Partículas com Deslocamento Aleatório (MLPDA) como uma alternativa ao método *box*. Seis formas para a função *kernel* são inseridas no MLPDA: gaussiana, Epanechnikov, bi-, tri-, quad- e quintweight. Testes para validar esse método e comparar suas seis formas mencionadas são conduzidos em uma baía hipotética quadrada com $5.000 \times 5.000m$ de extensão e forçada por uma onda de maré através do modelo hidrodinâmico DIVAST (bidimensional integrado na vertical). Nos testes, tanto o método *kernel*, quanto o *box*, são confrontados com uma curva de concentração analítica. Todas as formas do *kernel* alcançam resultados precisos de concentração com um número bem reduzido de partículas em comparação ao método *box*. Dentre as formas do *kernel*, a gaussiana se mostra ligeiramente mais vantajosa que as demais.

ABSTRACT --- An algorithm related to the kernel density method for calculating the concentration of a pollutant is implemented in a Lagrangian Model with Particle Tracking (MLPDA) as an alternative to the box method. Six forms for the kernel function are inserted in MLPDA: Gaussian, Epanechnikov, bi-, tri-, quad-and quintweight. Tests to validate this method and compare its six forms mentioned are conducted on a hypothetical bay with $5.000 \times 5.000m$ in length and forced by a tidal wave through the hydrodynamic model DIVAST (depth integrated velocity and solute transport). In testes, the kernel method and box method are compared with an analytical curve of concentration. All forms of the kernel show accurate results of concentration with a small number of particles compared to the method box. Among the forms of the kernel, the Gaussian is shown slightly more advantageous than the others.

Palavras chaves: *kernel* de densidade, modelagem lagrangiana de corpos d'água.

1) Mestrando em Engenharia Ambiental da UFES, Av. Fernando Ferrari, 514, 29075-910, Goiabeiras, Vitória-ES. E-mail: fngaze@gmail.com.

2) Professor Adjunto da UFES, CT, Av. Fernando Ferrari, 514, 29075-910, Goiabeiras, Vitória-ES. E-mail: juliotac@gmail.com.

3) Professor Adjunto da FAESA, Rodovia Serafim Derenzi, 3115, 29030-001, São Pedro, Vitória-ES. E-mail: maxsuelmrp@gmail.com.

1. INTRODUÇÃO

O Modelo Lagrangiano de Partículas com Deslocamento Aleatório (MLPDA) simula a dispersão de um dado poluente assumindo-o como um número finito de partículas, em que cada uma é tratada independentemente da outra. A base do MLPDA fundamenta-se na equação de Langevin (RODEAN, 1994):

$$dx_i = a_i(\mathbf{x}, t)dt + b_{ij}(\mathbf{x}, t)dW_j(t) \quad (1)$$

Onde $a_i(\mathbf{x}, t)dt$ é o termo de deriva (determinístico), $b_{ij}(\mathbf{x}, t)dW_j(t)$ o termo de difusão (estocástico) e $dW_j(t) = \xi_j dt$ um processo incremental de Wiener.

Na estimativa de concentração para um dado poluente, dois métodos são conhecidos: o método *box* de contagem de partículas e o método *kernel* de densidade.

O método *box*, segundo Haan (1999), consiste na contagem de todas as partículas em uma célula de grade no espaço (ver também MONFORTI *et al.*, 2006; VITALI *et al.*, 2006). A estimativa da concentração é, então, obtida multiplicando-se o número de partículas pelo valor de suas massas e dividindo essa massa total pelo volume do *box*.

Monforti *et al.* (2006) destacam que o método *box*, que se utiliza de uma grade euleriana, perde uma das principais vantagens da aproximação lagrangiana: a independência da grade computacional.

Por esse método, escolhas como a extensão e o centro do intervalo médio entre as áreas ou o volume no qual estão inseridas as partículas, são feitas. Haan (1999) observa que as diferenças entre estimativas da densidade da concentração com diferentes volumes amostrados são significantes. Fato esse que emprega ao método *box* certas incertezas.

Um meio de minimizar essas incertezas é aumentar o número de partículas simuladas. Borgas & Sawford (1994) assumem teoricamente que o número de partículas será alto o suficiente a medida que a escolha das dimensões do *box* de contagem das partículas seja negligenciada. Entretanto, há um forte obstáculo para simular esse número “alto o suficiente”: limitações impostas pela capacidade de armazenamento e velocidade de processamento computacional.

Frente às adversidades inerentes ao método *box*, entra em cena o método *kernel* de densidade, em que cada partícula gera uma distribuição de massa ao longo do domínio físico. A densidade da concentração, num ponto qualquer, terá a contribuição de todas as partículas que se encontrem no domínio físico. Segundo Vitali *et al.* (2006), o perfil de densidade é dado pela função

kernel, enquanto que a quantidade de massa “espalhada” depende do comprimento de banda (*bandwidth*) associado a cada partícula.

Lorimer (1986) destaca que a eficiência computacional é a propriedade mais importante do método *kernel*. Esse fato se deve, em teoria, pela razão do método *kernel* requerer uma menor quantidade de partículas com relação ao método *box* para reproduzir resultados de concentração com o mesmo nível de erro para ambos os métodos. A demanda computacional para o *kernel*, em consequência, pode ser satisfeita por um simples microcomputador doméstico.

Estimadores de densidade *kernel* têm sido estudados pelos estatísticos por aproximadamente trinta anos (detalhes em Tapia & Thompson, 1978), sendo introduzidos em simulações para fluidos compressíveis por Lucy (1977) e, subseqüentemente, por Gingold e Monaghan (1982). Esses autores concordam que os estimadores de densidade *kernel* mostram-se estatisticamente mais precisão do que o método por simples contagem de partículas em células (método *box*).

Trabalhos envolvendo o método *kernel* na modelagem de corpos d’água ainda são escassos. Exemplos, porém podem ser vistos em Gingold e Monaghan (1982) e Rajagopalan *et al.* (1997), que aplicaram e discutiram a teoria do método *kernel* em modelos de partículas hidrodinâmicos. Ulterior a eles, Ridge (2002), também desenvolveu esse método em um estudo que aborda a modelagem da dispersão de poluentes em águas costeiras. Além deles, Jones *et al.* (1996) focaram nos métodos de seleção do comprimento de banda para a estimativa de densidade.

O foco desse trabalho está, portanto, na implementação e validação do método *kernel* de densidade junto a um modelo lagrangiano aplicado a corpos hídricos.

2. ESTIMATIVA DA CONCENTRAÇÃO PELO *KERNEL* DE DENSIDADE

O estimador *kernel* de densidade para a concentração normalizada c de n partículas de mesma massa num dado ponto x é:

$$c(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (2)$$

onde h é o comprimento de banda do *kernel* e K é a função *kernel*, que satisfaz $K(x) \geq 0 \quad \forall x$, normalizada de modo que:

$$\int K(x)dx = 1 \quad (3)$$

fazendo de c uma distribuição de densidade ($\int c(x)dx = 1$). Um dos *kernels* mais amplamente empregados é o gaussiano (K_G):

$$K_G(r) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} r^T r\right) \quad (4)$$

Onde d representa a dimensão. Para modelos bidimensionais, $d = 2$.

Além do *kernel* gaussiano, outras formas serão investigadas:

$$K_a(x) = \begin{cases} C_{d,a} (1 - x^T x)^a & (x^T x < 1) \\ 0 & \text{caso contrário} \end{cases} \quad (5)$$

Na equação 5, $C_{d,a}$ são fatores de normalização garantindo que $\int K_a dx = 1$. Para valores de $a = 1$ até $a = 5$, esses *kernels* são chamados, respectivamente de Epanechnikov e *bi-*, *tri-*, *quad-* e *quintweight*.

Os valores de $C_{d,a}$, α e β são obtidos por (Haan, 1999):

$$C_{d,a} = \left\{ \pi^{d/2} \frac{\Gamma(a+1)}{\Gamma(a+1+d/2)} \right\}^{-1} = \frac{\prod_{i=1}^a (d+2i)}{c_d \cdot 2^a \cdot a!} \quad (6)$$

$$\alpha = C_{d,a} \frac{2^a a! c_d}{\prod_{i=1}^{a+1} (d+2i)} \quad (7)$$

$$\beta = C_{d,a}^2 \int (1 - x^T x)^{2a} dx = C_{d,a}^2 \left\{ \pi^{d/2} \frac{\Gamma(2a+1)}{\Gamma(2a+1+d/2)} \right\} = C_{d,a}^2 \frac{c_d 2^{2a} (2a)!}{\prod_{i=1}^{2a} (d+2i)} \quad (8)$$

As seis formas do *kernel* apresentadas anteriormente (gaussiano, Epanechnikov, *biweight*, *triweight*, *quadweight* e *quintweight*) são implementadas junto ao MLPDA.

3. VALIDAÇÃO DO MÉTODO *KERNEL*

3.1. As Bases dos Testes

Uma bateria de testes é conduzida a fim de avaliar a eficiência do cálculo de concentração das diversas formas do *kernel* em comparação ao método *box* e a uma curva de concentração analítica. Essa metodologia de validação parte de um princípio apresentado por autores como Haan (1999) e Ridge (2002), e consiste, basilarmente, de testar as formas do *kernel* com números distintos de partículas e encontrar qual delas fornecem curvas de concentração com baixos erros médios quadráticos em relação à uma curva gaussiana analítica, levando-se em conta o contrapeso do menor número possível de partículas.

3.2. O Domínio Computacional dos Testes

A Figura 1 expõe a área tomada para a condução dos testes. Trata-se de uma baía hipotética quadrangular de $5.000 \times 5.000 \text{ m}$. É ainda assumida uma profundidade constante de 10 m .

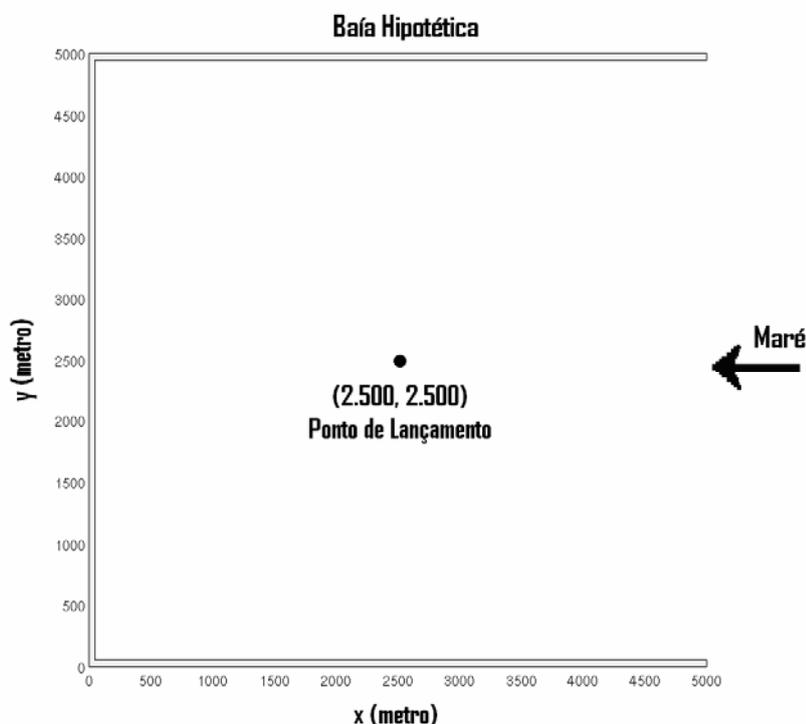


Figura 1. Baía hipotética quadrangular nos quais os testes de validação do método do *kernel* de densidade são realizados.

A hidrodinâmica da baía é simulada pelo modelo computacional DIVAST (*Depth Integrated Velocity and Solute Transport*). Desenvolvido pelo britânico Falconer (1976) e codificado em linguagem Fortran 77, esse modelo é empregado em estudos concernentes à hidrodinâmica e qualidade de água em regiões rasas. Do tipo bidimensional horizontal integrado na vertical (2DH), sua modelagem tem suporte nas equações de transporte de Reynolds com o uso do método numérico das diferenças finitas na solução das equações governantes do escoamento.

Para a região dos testes, adota-se um passo de tempo para a modelagem de 5 s . Por sua vez, o espaçamento das células da malha computacional é de $50 \times 50 \text{ m}$, que permite, tanto ao eixo x , quanto ao eixo y , um número total de 100 células. A hidrodinâmica da baía é estimulada por uma onda de maré de período igual a $12,4 \text{ h}$ e amplitude de 70 cm que alcança seu interior pelo contorno aberto. Essa modelagem hidrodinâmica é realizada pela partida a frio, onde assume inicialmente velocidade nula ao longo do domínio computacional e elevação da água a partir de seu nível

máximo (preamar). Os resultados dos campos de velocidade e da distribuição dos valores do coeficiente de dispersão são armazenados de hora em hora. A forçante devida ao vento não é considerada.

Partimos, agora, ao modelo de trajetória de partículas (MLPDA). Para todos os processamentos com as distintas formas do *kernel* e os diferentes números de partículas, o MLPDA estabelece as seguintes características: as partículas são lançadas instantaneamente após $24h$ no ponto central do domínio (2.500×2.500). O passo de tempo lagrangiano é de $0,5s$. A massa total do lançamento é de $1Kg$, sendo as curvas de concentração analisadas em $74h$, ou seja, $50h$ após o lançamento das partículas. Pela Figura 2, é possível observar um padrão aproximadamente gaussiano das posições das partículas após o tempo esse tempo de $50h$.

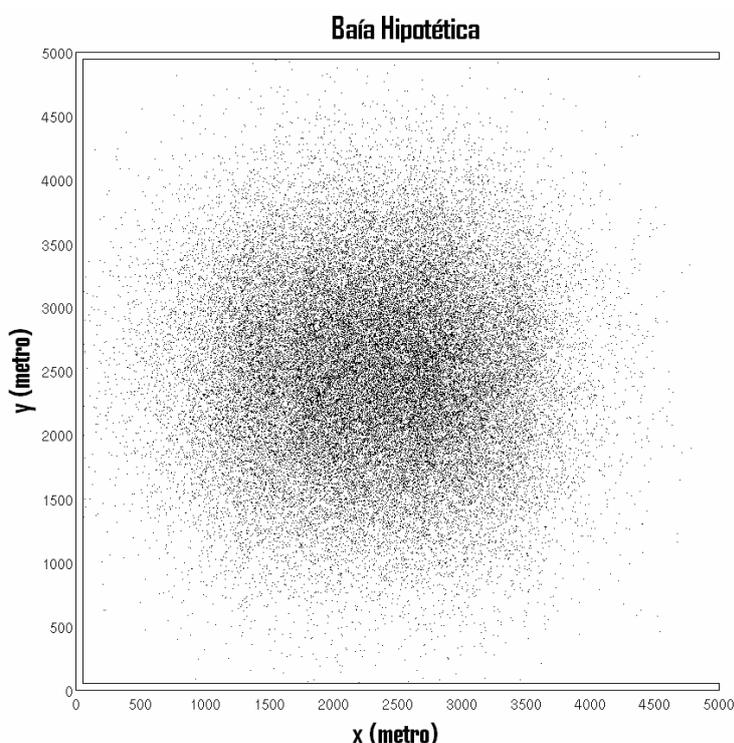


Figura 2. Posição de 50 mil partículas calculadas após 50h do lançamento.

Ainda sobre as curvas de concentração, elas são construídas em um plano transversal em $y = 2.500m$ e variando o eixo x de um extremo ao outro (ponto 0 ao ponto $5.000m$).

Consta-se que todos os processamentos são conduzidos num microcomputador com processador Intel Core 2 Duo, 2 Gigabytes de memória RAM e sistema operacional Linux openSUSE 10,3.

3.3. Os Resultados dos Testes

Os primeiros gráficos (Figuras 3 a 5) contam o comportamento das curvas de concentração calculadas pelo método *box* para distintos números de partículas lançadas:

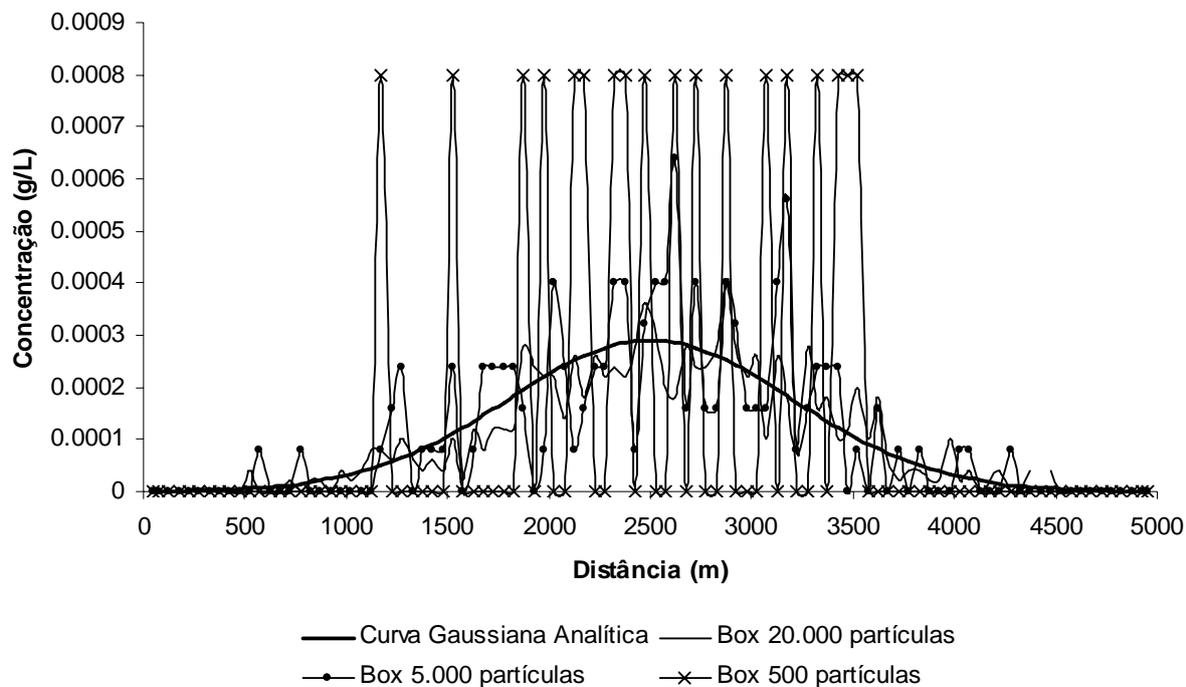


Figura 3. Concentrações vs. distância calculadas pelo método *box* para lançamentos com 20 mil, 5 mil e 500 partículas após 50 h do lançamento das partículas.

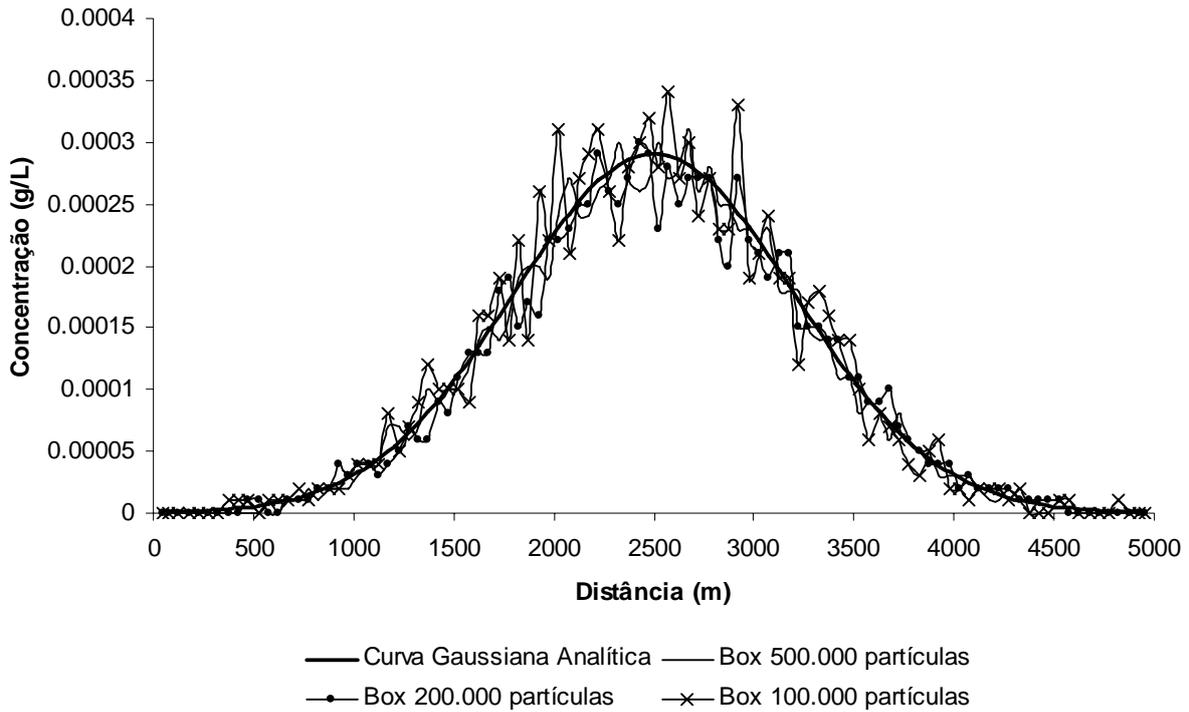


Figura 4. Concentrações vs. distância calculadas pelo método *box* para lançamentos com 500, 200 e 100 mil partículas após 50h do lançamento das partículas.

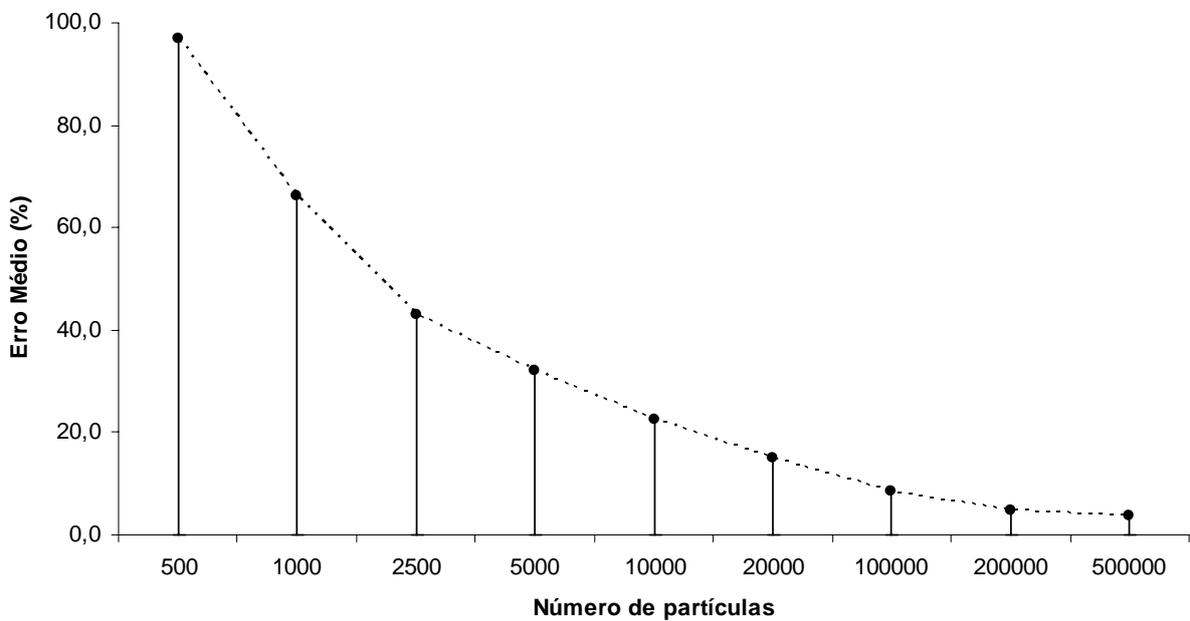


Figura 5. Percentual do erro médio quadrático para o método *box* com diferentes números de partículas em comparação à concentração da curva gaussiana analítica.

Os produtos dos testes oriundos dos cálculos da concentração pelas diversas formas assumidas pelo *kernel* estão compilados abaixo:

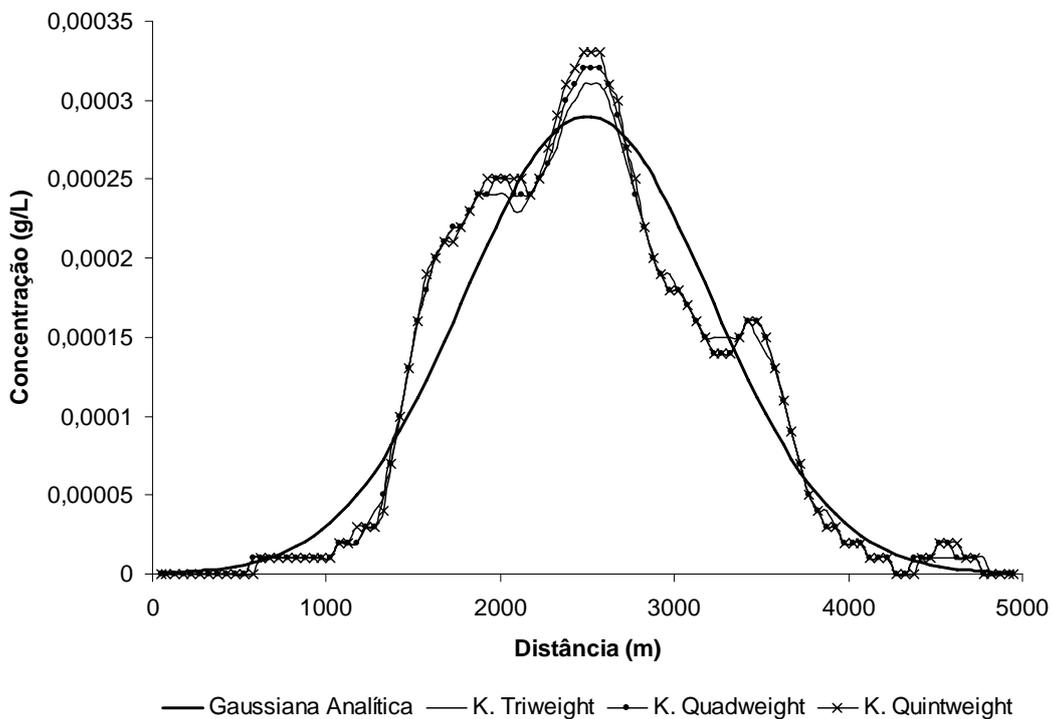
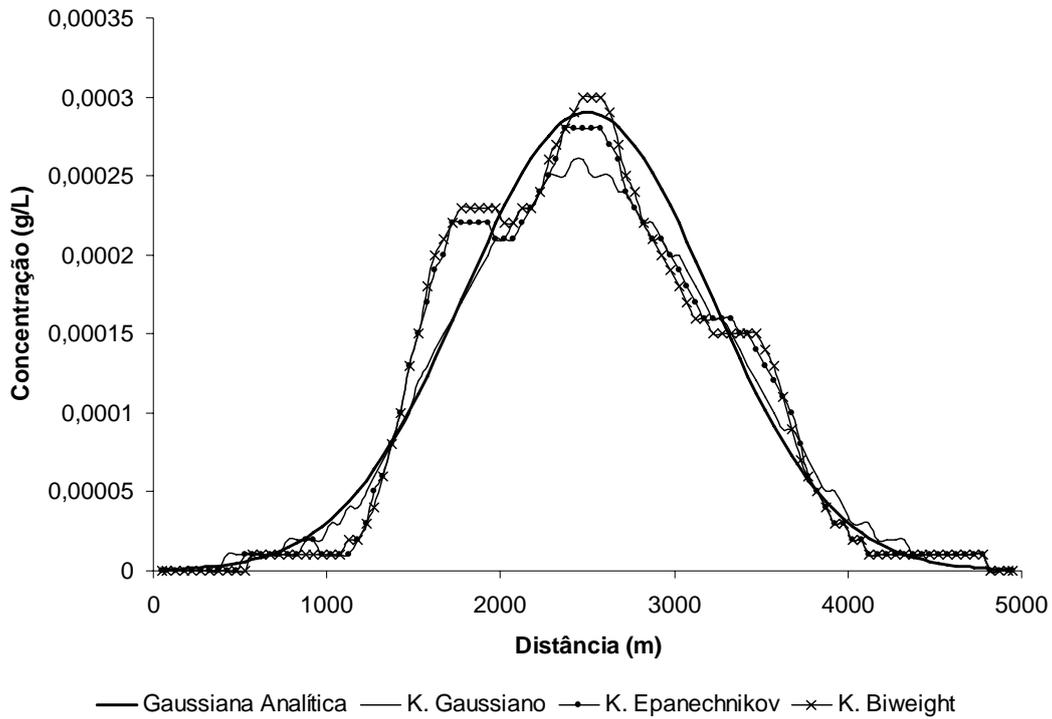


Figura 6. Gráficos das concentrações vs. distância calculadas pelas seis formas do *kernel* para lançamentos com 500 partículas após 50 h do lançamento das partículas.

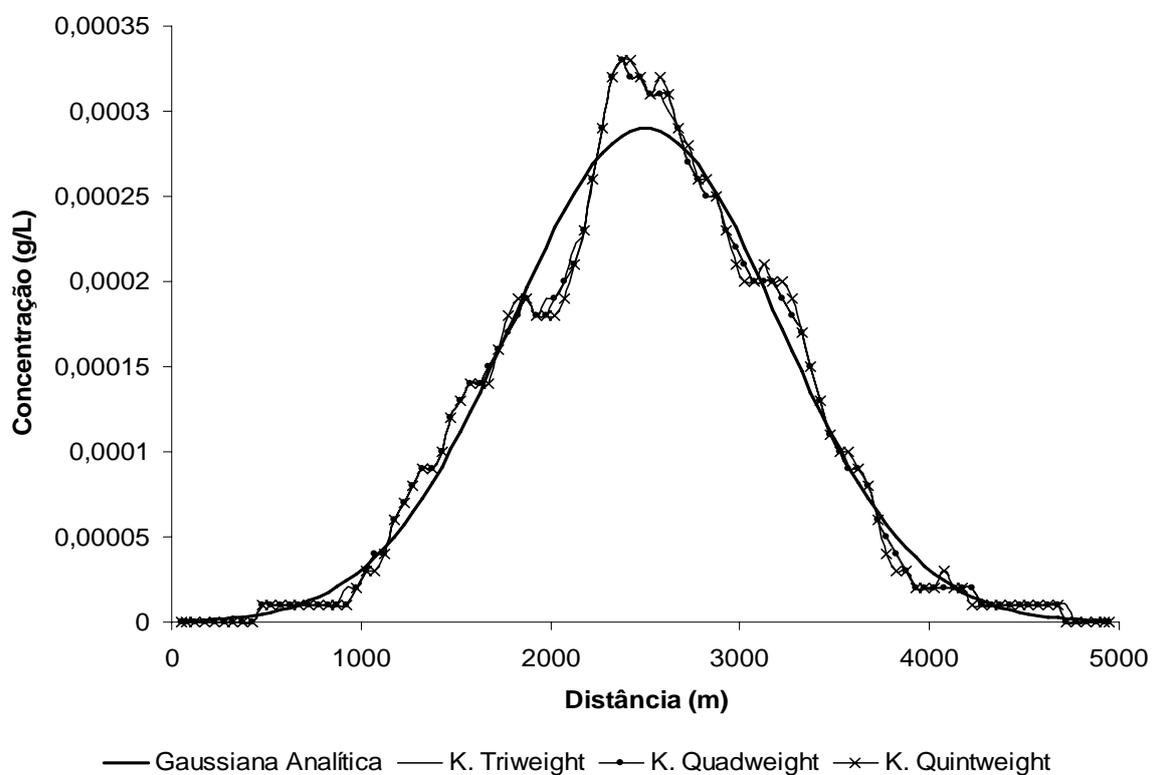
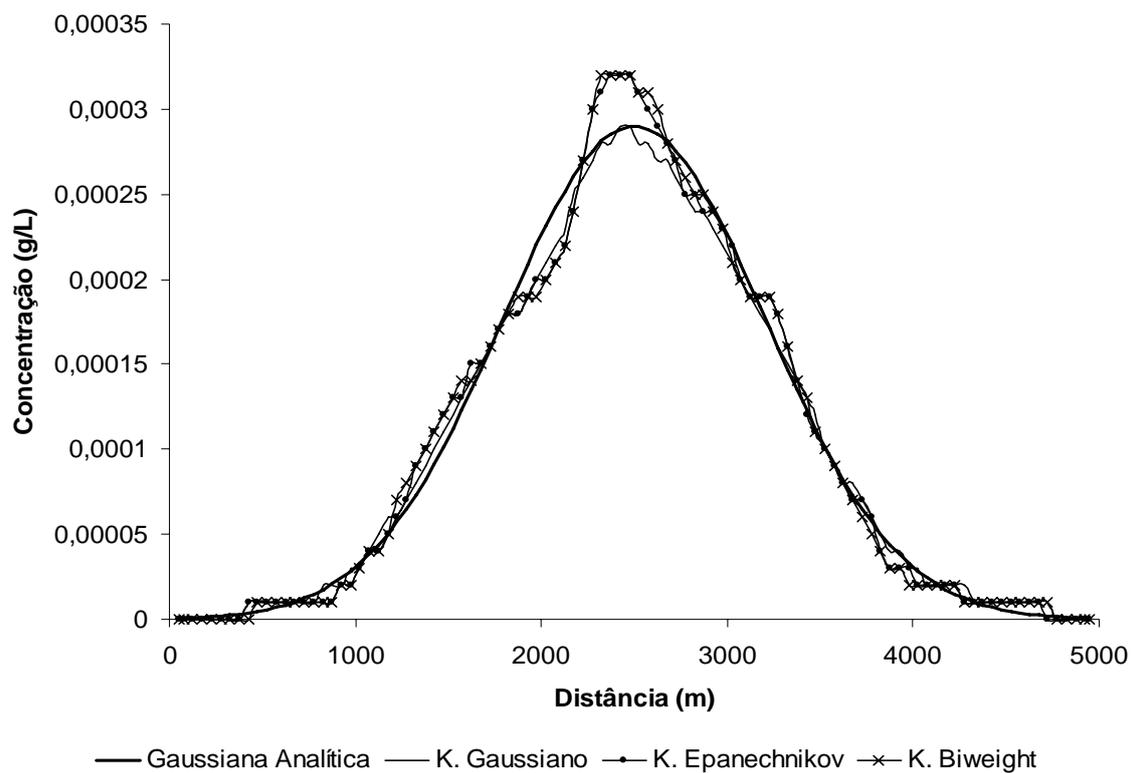


Figura 7. Gráficos das concentrações vs. distância calculadas pelas seis formas do *kernel* para lançamentos com 5 mil partículas após 50h do lançamento das partículas.

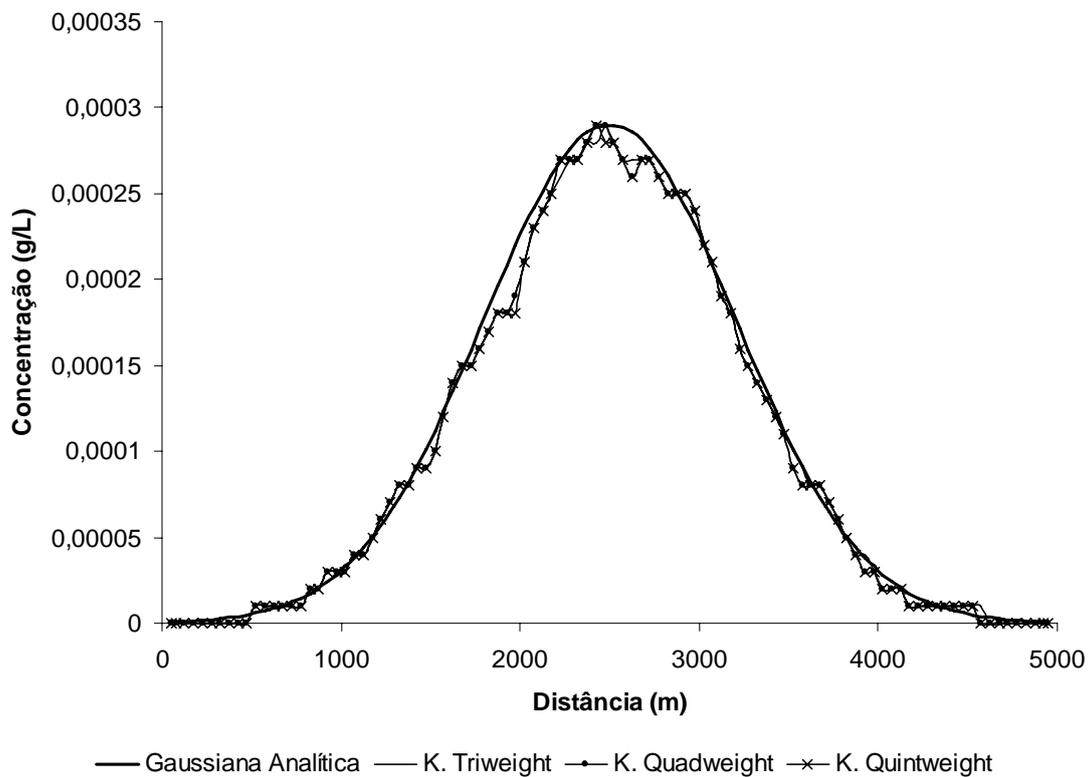
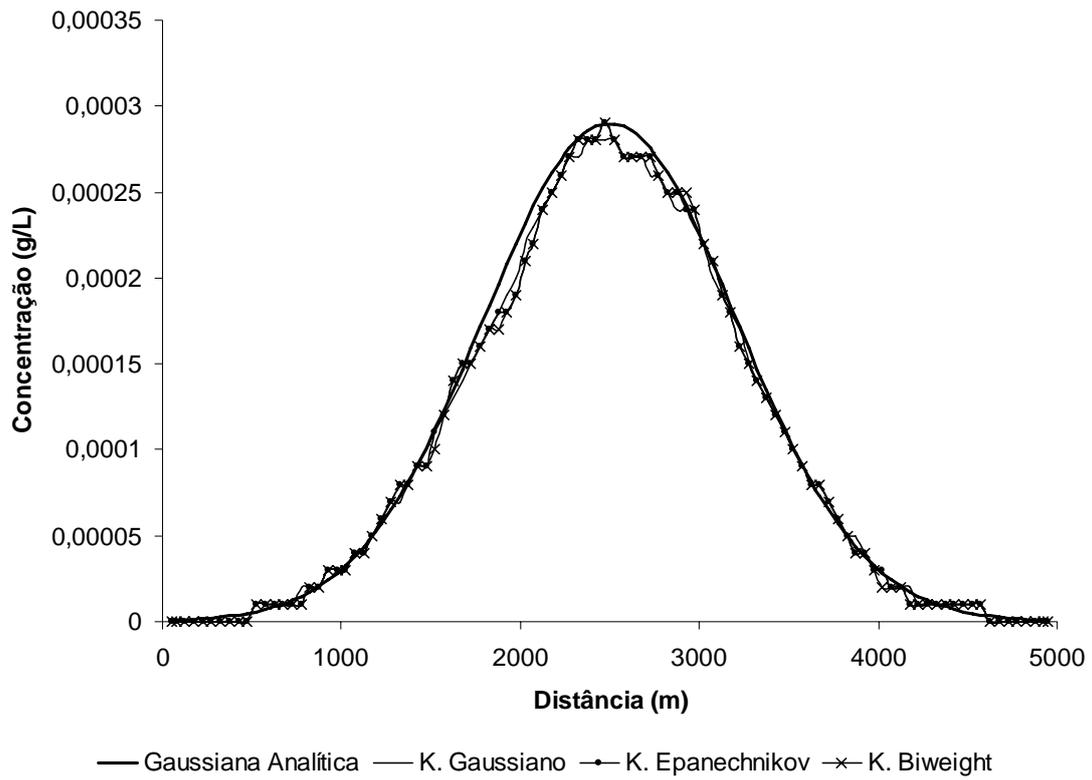


Figura 8. Gráficos das concentrações vs. distância calculadas pelas seis formas do *kernel* para lançamentos com 50 mil partículas.

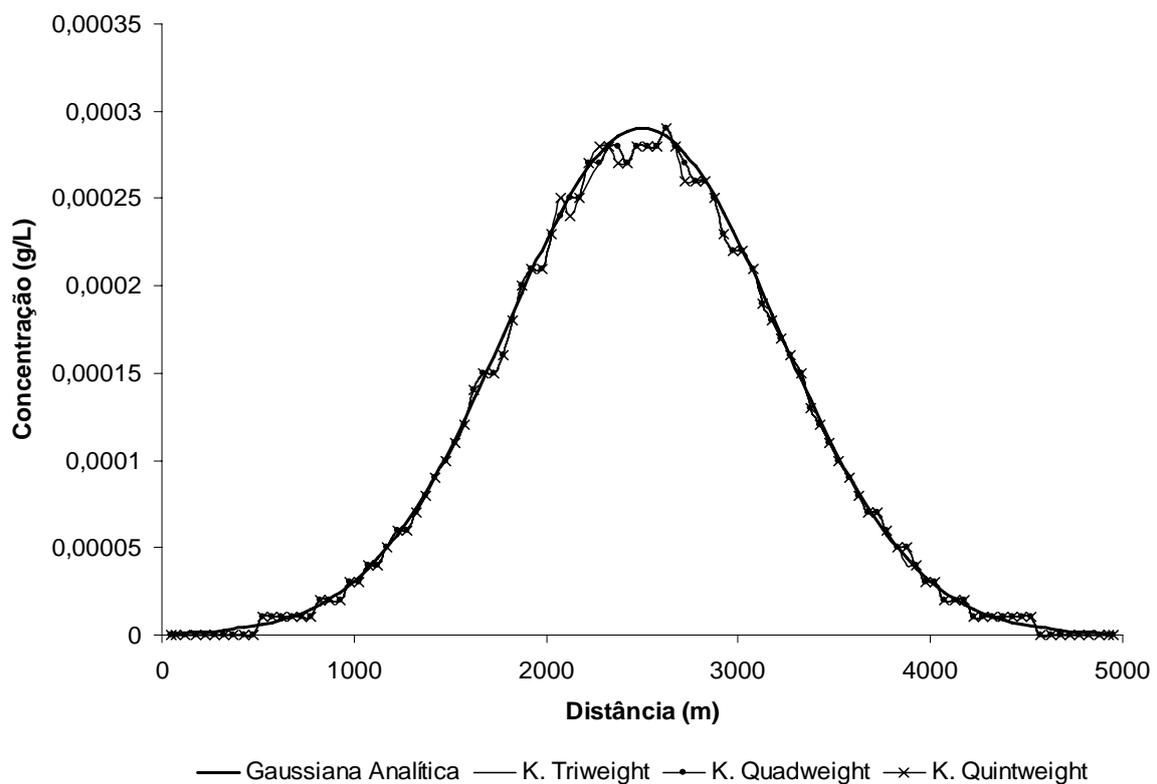
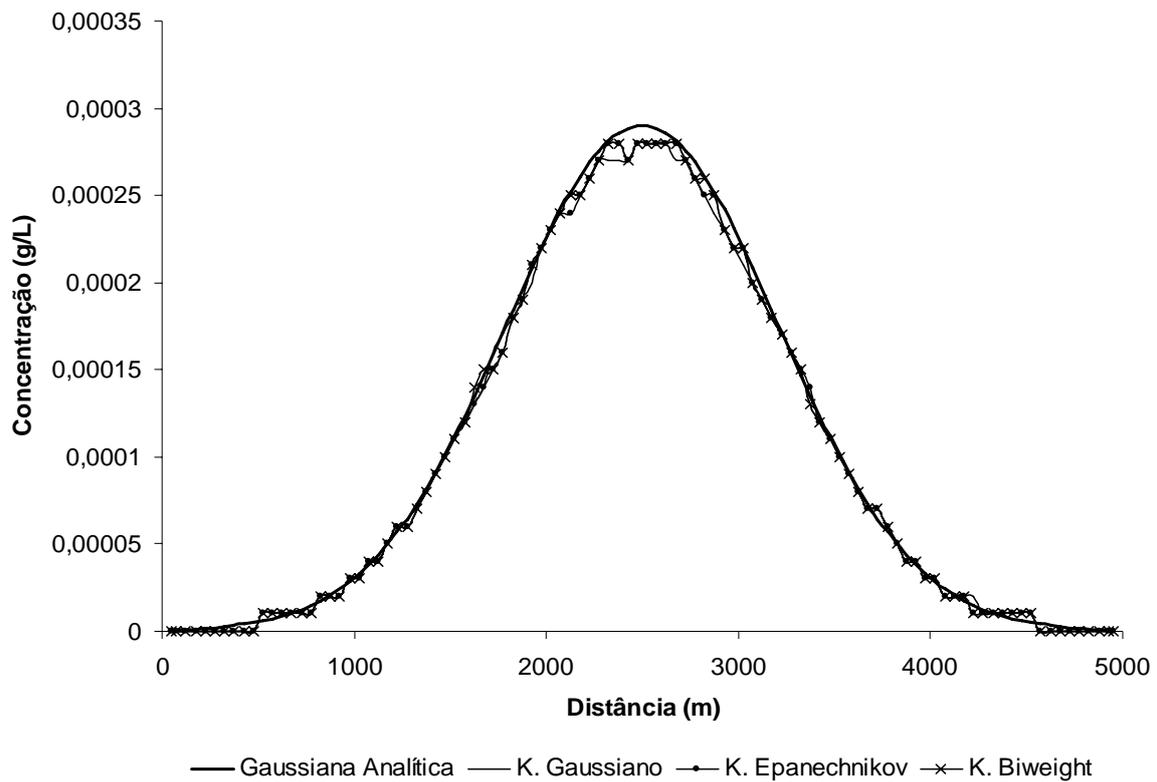


Figura 9. Concentrações vs. distância calculadas por seis formas admitidas pelo *kernel* para lançamentos com 500 mil partículas.

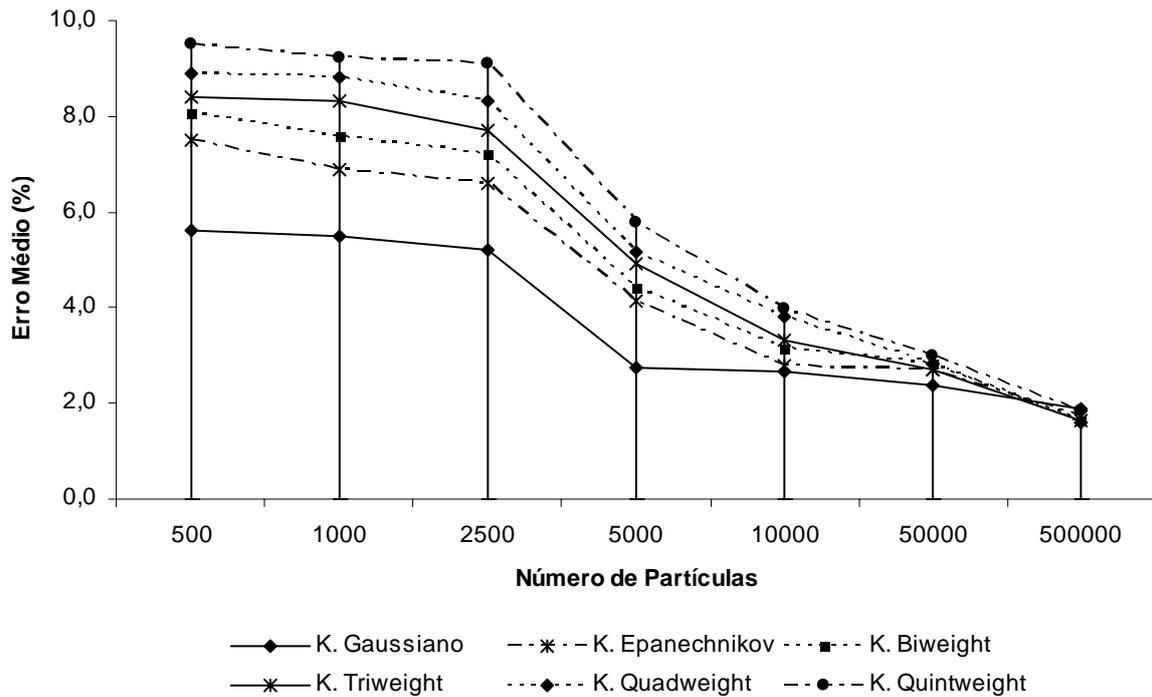


Figura 10. Percentual do erro médio quadrático para os diferentes *kernels* com distintos números de partículas em comparação à curva gaussiana analítica.

4. DISCUSSÃO

Conforme se espera, o acréscimo do número de partículas lançadas no domínio computacional redundam em maior precisão dos cálculos de concentração com precedente declínio do erro médio quadrático.

O método *box* exibe curvas acentuadamente irregulares quando o lançamento é realizado com números de partículas inferiores a 100 mil (Figura 3). Os resultados são significantes erros médios, que abordam próximos dos 100% em processamentos conjugados a 500 partículas (Figura 5). Sem especificação de suas formas, o método *kernel* (Figuras 6 a 9), embora produza certas irregularidades para baixos números de partículas, não geram erros que alcançam os 10% em lançamentos com 500 partículas (Figura 10).

500 mil partículas é o valor que determina resultados satisfatórios em processamentos com o método *box* (Figura 4). No caso, o erro médio quadrático oscila em torno de 3,5% (Figura 6). O décimo desse valor (50 mil) é o suficiente para o método *kernel* produzir erros que não cruzam a fronteira dos 3% (Figura 10).

No embate, o *kernel* leva ampla vantagem sobre o método de contagem de partículas em células, visto que, um número significativamente menor de partículas resulta em grande otimização computacional, que se traduz como baixos tempos de processamento (Gingold & Monaghan, 1982; Lorimer, 1986; Haan, 1999). As modelagens geridas com 500 mil partículas demandam cerca de 23 min por saída de dados (1h do processo modelado), e da ordem de 3 min para 50 mil partículas. Salienta-se que não existem discrepâncias expressivas nesses tempos quando, com uma quantidade fixa de partículas, a comparação é do *kernel* (analisando todas as formas) com o *box*.

A confrontação das formas do *kernel* é definida a favor da gaussiana. Desde os testes com baixas quantidades de partículas, chegando aos com números mais elevados (Figuras 6 a 9), provém da função gaussiana os menores erros médios quadráticos (Figura 10).

Além das trajetórias das partículas no domínio computacional apontarem uma distribuição aproximadamente normal (Figura 2), outro agente que é acusado a essa inclinação do modelo lagrangiano responder melhor à forma gaussiana é a opção do algoritmo de geradores de números pseudo-aleatórios, o GASDEV, cujo algoritmo gera variáveis com distribuição gaussiana.

A despeito da vantagem do *kernel* gaussiano, as demais funções não expõem um desempenho ruim em processamentos a partir de 50 mil partículas. Nesses valores, é factível se trabalhar com confiabilidade com qualquer forma mencionada. Os testes ainda autenticam os conduzidos por Haan (1999), que apontam uma discrepância desprezível entre os erros médios das funções do *kernel* quando se trabalha com 500 mil partículas. O máximo não ultrapassou o valor de 0,2% (Figura 10).

AGRADECIMENTOS

À Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) pelo apoio financeiro à pesquisa;

Ao Laboratório de Simulação em Escoamentos com Superfície Livre (LABESUL) pelo espaço físico.

BIBLIOGRAFIA

BORGAS, M S. & SAWFORD, B. L.. **A Family of Stochastic Models for Two-particle Dispersion in Isotropic Homogeneous Stationary Turbulence**. Journal of Fluid Mechanics, v. 279, p. 69 – 99, 1994.

FALCONER, R.A. **Mathematical Modelling of Jet-Forced Circulation in Reservoirs and Harbours**. Tese. Dept. of Civil Engineering of Imperial College, London, 1976.

- GINGOLD, R. A.; MONAGHAN, J. J.. **Kernel Estimates as a Basis for General Particle Methods in Hydrodynamics**. Journal of Computational Physics, v. 46, p. 429 – 453, 1982.
- HAAN, P. de. **On the Use of Density Kernels for Concentration Estimations Within Particle and Puff Dispersion Models**. Atmospheric Environment, v. 33, p. 2007 – 2021, 1999.
- JONES, M. C. *et al.* **A Brief Survey of Bandwidth Selection for Density Estimation**. Journal of the American Statistical Association, v. 91, p. 401 – 407, 1996.
- LORIMER, G... **The Kernel Method for Air Quality Modelling – I. Mathematical Foundation**. Atmospheric Environment, v. 20, p. 1447 – 1452, 1986.
- LUCY, L. B.. **A Numerical Approach to Testing the Fission Hypothesis**. Astronomical Journal, v. 82, p. 1013 – 1024, 1977.
- MONFORTI, F. *et al.* **Testing Kernel Density Reconstruction for Lagrangian Photochemical Modelling**. Atmospheric Environment, v. 40, p. 7770 – 7785, 2006.
- RAJAGOPALAN, B. *et al.* **Evaluation of Kernel Density Estimation Methods for Daily Precipitation Resampling**. Stochastic Hydrology and Hydraulics. Springer-Verlag, 1997.
- RIDGE, M. M.. **Three-Dimensional Simulation of Pollutant Dispersion in Coastal Waters**. Tese de Doutorado. Universitat Politècnica de Catalunya, 2002.
- RODEAN, H.C.. **Notes on the Langevin Model for Turbulent Diffusion of “Marked” Particles**. Lawrence Livermore National Laboratory, EUA, 1994.
- TAPIA, R. A.; THOMPSON, J. R.. **Non-parametric Probability Density Estimation**. Johns Hopkins University Press, Baltimore, 1978.