DEFENCE **R&D** DÉFENSE

The Rapid Evaluation of Mean Concentration Fields in Lagrangian Stochastic Modelling Using a Density Kernel Estimator

Shao, Y. and Yee. E.
Defence R&D Canada – Suffield

Technical Report

DRDC Suffield TR 2004-186

October 2004

Canada

20050112 102

The Rapid Evaluation of Mean Concentration Fields in Lagrangian Stochastic Modelling Using a Density Kernel Estimator

Shao, Y. and Yee. E.
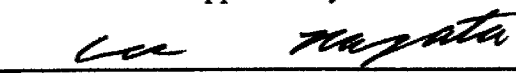Defence R&D Canada – Suffield
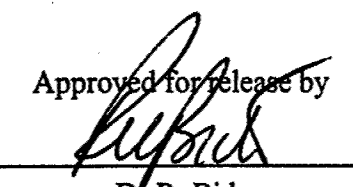
Defence R&D Canada – Suffield

Author

Y. Shao

Approved by

Dr J. Lavigne
Head, Chemical and Biological Section

Approved for release by

Dr R. Bide
Chair, DRDC Suffield DRP

Abstract

Lagrangian Stochastic (LS) particle models have proven to be a useful computational tool for the description and prediction of dispersion of pollutant releases in complex meteorological situations (e.g., space- and time-varying situations pertaining to complex flow and turbulence). However, simulating the emitted pollutant by following the trajectories of many "marked" fluid elements released from the source distribution brings up the difficulty of the correct estimation of the mean concentration of the dispersing pollutant from the particle trajectory information. Recently, the density kernel estimation method has been proposed and applied successfully to estimate mean concentrations from Lagrangian Stochastic particle models. However, the computational effort needed by this method increases as N^2 (assuming the number of receptor locations N_r at which the concentration is required is comparable to the number of fluid particles N_p used in the trajectory simulation, so $N_r \approx N_p \sim N$) and, in consequence, the method has not been widely used because of the significant computer resources required. Here, we describe a novel algorithm for calculating the kernel estimate of the mean concentration field whose computational complexity scales only as N. The technique uses a tesselation (subdivision) of space in cubic cells of side length h (where h is the bandwidth of the kernel function), and then associates a linked-list data structure with each cell that is used as a bookkeeping device to keep track of the "marked" fluid particles in that cell. The fast approach developed here has been verified by comparing results with the direct implementation of the kernel estimator and with the conventional box-counting estimator for the mean concentration field.

Résumé

On a prouvé que les modèles de particules langrangiens stochastiques (LS) sont des outils utiles de calcul pour décrire et prédire la dispersion des émissions de polluants dans des situations météorologiques complexes (par ex. : des situations dans l'espace et le temps qui varient selon des flux et des turbulences complexes). Simuler le polluant émis en suivant les trajectoires de beaucoup d'éléments de fluides « marqués » qui ont été émis à la répartition des sources, augmente cependant la difficulté d'estimer correctement la concentration moyenne du polluant dispersé, à partir de l'information sur la trajectoire de la particule. La méthode d'estimation à partir du noyau de densité a été récemment proposée et appliquée avec succès pour estimer les concentrations moyennes des modèles de particules langrangiens stochastiques. Cependant, l'effort de calcul requis par cette méthode augmente par un facteur N^2 (dans l'hypothèse où le nombre de lieux récepteurs N_r, auxquels la concentration est requise est comparable au nombre de particules de fluides N_r, utilisé dans la simulation de la trajectoire, tel que $N_r \approx N_p \sim N$) et, par conséquent, cette méthode, exigeant des ressources informatiques très importantes, n'a pas été très utilisée. Nous décrivons ici, un nouvel algorithme qui calcule l'estimation du noyau de densité du champ de la concentration moyenne et dont la complexité des calculs est réduite à N. La technique utilise une tessellation (subdivision) de l'espace en cellules cubiques dont la longueur des côtés h (h étant la largeur de bande de la fonction du noyau) et puis associe une structure de données en chaîne avec chaque cellule; cette structure, étant utilisée comme engin de comptabilité, suit les particules de fluides « marquées » dans cette cellule. Cette méthode rapide, mise au point ici, a été vérifiée en comparant des résultats avec l'implémentation directe de l'estimateur de noyau et avec l'estimateur classique de « comptage de boîtes » pour le champ de concentration moyenne.

Executive summary

Background

There are, quite rightly, growing concerns world-wide about the dangers, both actual and potential of the use (release) of chemical and biological warfare (CBW) agents into the atmosphere. The increased awareness and importance accorded by the public world-wide and their governments to maintain appropriate defences against CBW agents in the environment, the prediction of casualties and human performance degradation resulting from such releases, and the development of operational procedures and regulations to control, mitigate, and monitor the fate of CBW agents in the atmosphere will require mathematical modelling of atmospheric dispersion of these agents. This modelling is necessary in order to answer political and planning questions such as: what is the extent of the hazard zone corresponding to the release of a CBW agent and what is the effect of this release on exposed personnel within the hazard region; and what contingency measures would be needed to deal with a potential release of a CBW agent. Accurate prediction of the dispersion of contaminants released into the complex atmospheric boundary layer is a particularly challenging problem. Compared to conventional techniques for the calculation of dispersion (e.g., similarity theory, statistical theory, or eddy-diffusivity methods), Lagrangian Stochastic (LS) models of particle motions have proven to be a successful and flexible tool in the description and prediction of contaminant dispersion in complex meteorological situations (incorporating in a natural manner the effects of inhomogeneities, unsteadiness and/or non-Gaussianity in turbulent velocity distributions associated with the necessarily complex flow and turbulence encountered in the real-world atmosphere).

The major disadvantage of the LS modelling approach is the need for large computational resources to run the model for typical applications in order to obtain statistically meaningful solutions of contaminant dispersion and mean concentrations for sources in typical CBW agent releases. To facilitate the application of state-of-the-art LS models for CBW agent dispersion predictions, we developed a computationally efficient algorithm for mean concentration estimation based on the density kernel method that can be used in conjunction with LS models of dispersion.

Principal results

Lagrangian Stochastic particle models have proven to be a useful computational tool for the description and prediction of dispersion of pollutant releases in complex meteorological situations (e.g., space- and time-varying situations pertaining to complex flow and turbulence). However, simulating the emitted pollutant by following the trajectories of many "marked" fluid elements released from the source distribution brings up the difficulty of the correct estimation of the mean concentration of the dispersing pollutant from the particle trajectory information. Recently, the density kernel estimation method has been proposed and applied successfully to estimate mean concentrations from LS particle models. However, the computational effort needed by this method increases as N^2 (assuming the number of receptor locations N_r at which the concentration is required is comparable to the number of fluid particles N_p used in the trajectory simulation, so $N_r \approx N_p \sim N$) and, in consequence, the method

has not been widely used because of the significant computer resources required. Here, we describe a novel algorithm for calculating the kernel estimate of the mean concentration field whose computational complexity scales only as N. The technique uses a tesselation (subdivision) of space in cubic cells of side length h (where h is the bandwidth of the kernel function), and then associates a linked-list data structure with each cell that is used as a bookkeeping device to keep track of the "marked" fluid particles in that cell. The fast approach developed here has been verified by comparing results with the direct implementation of the kernel estimator and with the conventional box-counting estimator for the mean concentration field.

Significance of results

The development and implementation of a fast linked list based algorithm for a density kernel estimate of the mean concentration field has been described, allowing the user to calculate a set of kernel estimates of concentration at N_r receptor locations in order $O(N_r)$ work as opposed to order $O(N_r^2)$ effort (assuming the typical situation where the number of fluid particles N_p used in the trajectory-simulation satisfies $N_p \geq N_r$). The algorithm makes feasible the use of LS models on a personal computer for the routine calculation of dispersion associated with complex CBW agent release scenarios into the atmosphere.

Future work

There are several ways in which the code for the fast linked list based algorithm for the density kernel estimator of the mean concentration field can be generalized and/or made more efficient. The linked list data structure used here is appropriate for the case where the bandwidth of the density kernel associated with each particle is fixed. In principle, the bandwidth h need not be constant and for many applications it may be useful to allow a variable smoothing length that is dynamically adapted so that the number of neighboring fluid particles within the support of a kernel function centered on any particular particle remains constant. The generalization of the proposed fast kernel estimator for variable kernel bandwidth h will probably require that the linked list data structure used here be replaced by a hierarchy tree data structure that can be adopted to suit the needs of a variable bandwidth. In the application of the fast kernel estimator, the code can be made even more efficient by pre-computing the values of the kernel function for a large number of representative points and storing the results in a pre-computed kernel function value look-up table. Finally, it may well be worth exploring the parallelization of the code so that it can be executed on computers with highly parallel architectures (e.g., Beowulf clusters).

Shao, Y. and Yee, E. (2004). The Rapid Evaluation of Mean Concentration Fields in Lagrangian Stochastic Modelling Using a Density Kernel Estimator. (DRDC Suffield TR 2004-186). Defence R&D Canada – Suffield.

Sommaire

Contexte

Il existe, à juste titre, des inquiétudes croissantes au niveau mondial, en ce qui concerne les dangers, à la fois actuels et potentiels, d'utilisation (émission) d'agents chimiques et biologiques de guerre (CBW), dans l'atmosphère. Cette conscience accrue et l'importance accordée par le public et par les gouvernements, au niveau mondial, le maintien des défenses appropriées contre les agents CBW dans l'environnement, la prévision de victimes, la dégradation des performances humaines résultant de telles émissions et l'élaboration de procédures opérationnelles et de règlements pour contrôler, limiter et surveiller le sort des agents CBW dans l'atmosphère, exigeront la modélisation mathématique de la dispersion de ces agents dans l'atmosphère. Cette modélisation est nécessaire pour faire face aux problèmes politiques et de planification tels que : l'étendue de la zone dangereuse résultant de l'émission d'un agent CBW et l'effet d'une telle émission sur le personnel qui a été exposé à l'intérieur de la région dangereuse ; et les mesures d'urgence requises pour gérer l'émission potentielle d'un agent CBW. Prédire avec exactitude la dispersion de ces contaminants émis dans le couche limite atmosphérique est très problématique. Comparés aux techniques classiques de calcul de la dispersion (par ex. : théorie de la similitude, théorie statistique ou les méthodes de diffusivité des remous), les mouvements des modèles de particules lagrangiens stochastiques (LS) se sont prouvés être des outils souples qui réussissent à décrire et à prédire la dispersion de contaminants dans des situations météorologiques complexes (incorporant de manière naturelle les effets d'éléments non homogènes, le manque de régularité et /ou les distributions non gaussiennes de vitesse limite de régime turbulent associées avec les flux et les turbulences complexes qui sont nécessairement rencontrées dans la réalité de l'atmosphère).

L'inconvénient principal de cette méthode de modélisation LS est que cette dernière nécessite des ressources très importantes en puissance de calcul pour que ce modèle produise des solutions aux applications caractéristiques qui soient statistiquement valables en ce qui concerne la dispersion de contaminants et les concentrations moyennes des sources, lors d'émissions caractéristiques d'agents CBW. Un algorithme de traitement efficace estimant la concentration moyenne basée sur la méthode de la densité du noyau a été développé pour faciliter l'application des ces modèles LS, les plus récents de la technique, aux prédictions de dispersion d'agents CBW et peuvent être utilisés en conjonction avec les modèles LS de dispersion.

Les résultats principaux

On a prouvé que les modèles de particules langrangiens stochastiques sont des outils utiles de calcul pour décrire et prédire la dispersion des émissions de polluants dans des situations météorologiques complexes (par ex. : des situations dans l'espace et le temps qui varient selon des flux et des turbulences complexes). Simuler le polluant émis en suivant les trajectoires de beaucoup d'éléments de fluides « marqués » qui ont été émis à la répartition des sources, augmente cependant la difficulté d'estimer correctement la concentration moyenne du polluant dispersé, à partir de l'information sur la trajectoire de la particule. La méthode d'estimation à partir du noyau de densité a été récemment proposée et appliquée avec succès pour estimer les concentrations moyennes des modèles de particules langrangiens stochastiques. Cependant, l'effort de calcul requis par cette méthode augmente par un facteur N^2 (dans l'hypothèse où le nombre de lieux récepteurs N_r auxquels la concentration est requise est comparable au nombre de particules de fluides N_r utilisé dans la simulation de la trajectoire, tel que $N_r \approx N_p \sim N$) et, par

conséquent, cette méthode, exigeant des ressources informatiques très importantes, n'a pas été très utilisée. Nous décrivons ici, un nouvel algorithme qui calcule l'estimation du noyau du champ de la concentration moyenne et dont la complexité des calculs est réduite à N. La technique utilise une tessellation (subdivision) de l'espace en cellules cubiques dont la longueur des côtés h (h étant la largeur de bande de la fonction du noyau) et puis associe une structure de données en liste chaînée avec chaque cellule; cette structure étant utilisée comme engin de comptabilité suivant les particules fluides « marquées » dans cette cellule. Cette méthode rapide, mise au point ici, a été vérifiée en comparant des résultats avec l'implémentation directe de l'estimateur de noyau et avec l'estimateur classique de « comptage de boîtes », pour le champ de concentration moyenne.

La portée des résultats

On décrit le développement et l'implémentation d'un algorithme à base de données en liste chaînée rapide qui estime le noyau de la densité dans le champ de concentration moyenne, ce qui permet à l'utilisateur de calculer un ensemble d'estimations du noyau de la concentration aux lieux récepteurs N_r, d'ordre $O(N_r)$ du travail contrairement à l'ordre $O(N_r^2)$ de l'effort). (Dans l'hypothèse d'une situation caractéristique où le nombre de particules fluides N_p utilisé dans la simulation de la trajectoire correspond à $N_p \geq N_r$). L'algorithme rend l'utilisation des modèles LS faisable avec un ordinateur personnel pour les calculs routiniers de la dispersion, associés aux scénarios complexes d'émissions d'agents CBW, dans l'atmosphère.

Les travaux futurs

Il existe plusieurs façons de généraliser le code des algorithmes de données à base de liste chaînée rapide qui vise à estimer la densité du noyau, dans le champ de concentration moyenne, et / ou de rendre ce code plus efficace. La structure des données à base de liste chaînée utilisée ici est appropriée dans le cas où la largeur de bande de la densité du noyau associée à chaque particule est fixe. En principe, la largeur de bande h n'a pas besoin d'être constante et dans beaucoup d'applications, il peut suffir d'allouer une longueur variable souple pouvant dynamiquement s'adapter, en sorte que le nombre de particules fluides avoisinantes, à l'intérieur du support de la fonction du noyau centrée sur une particule donnée, reste constant. La généralisation de l'estimateur rapide de noyau qui est proposé pour la largeur de bande h variable exigera probablement que la structure des données à base de liste chaînée utilisée ici soit remplacée par une structure de donnée, un organigramme hiérarchique, pouvant être adopté pour satisfaire les besoins d'une largeur de bande variable. Dans l'application de l'estimateur de noyau rapide, le code peut être rendu encore plus efficace en pré calculant les valeurs de la fonction du noyau pour un grand nombre de points représentatifs et en sauvegardant les résultats dans une table de recherche de la valeur de la fonction pré calculée des noyaux. Il pourrait enfin être intéressant d'explorer la parallélisation du code de manière à ce qu'il puisse être traité avec des ordinateurs possédant des architectures hautement parallèles (par ex. agrégats Beowulf).

Shao, Y. and Yee, E. (2004). The Rapid Evaluation of Mean Concentration Fields in Lagrangian Stochastic Modelling Using a Density Kernel Estimator. (DRDC Suffield TR 2004-186). R & D pour la défense Canada – Suffield.

Table of contents

List of figures

Introduction

The prediction of the dispersion of contaminants released into various turbulent flows is of practical importance in a number of applications ranging from industrial mixing problems to the analysis of nuisance and hazard in air quality, pollution and combustion problems involving the release of toxic or flammable materials. In consequence, much effort has been expended on the development of models to predict mean concentrations of contaminants for a given emission-source distribution. However, the accurate prediction of the transport and diffusion of passive scalars released in a spatially inhomogeneous and temporally non-stationary turbulent flow (e.g., inhomogeneous and non-stationary fields of mean wind speed and turbulence over complex terrain in the atmospheric boundary layer) is a challenging problem. For these types of turbulent flows, a Lagrangian Stochastic (LS) model describing the paths or trajectories of "marked" fluid elements offers great potential for studying contaminant dispersion since they can model properly inhomogeneities, non-steadiness, or non-Gaussianity in the background velocity distribution (viz., LS models of particle motion are a particularly attractive option for the simulation of dispersion in complex flows).

In a LS model, the mean concentration of a passive scalar transported by a turbulent flow field (whose velocity statistics are assumed to be known or prescribed *a priori*) is obtained from the probability distribution of the displacements of independent "marked" fluid elements (or particles of tracer which should not be viewed as physical particles, but rather as markers keeping track of the effects of fluid convection) released from the source distribution of the scalar. Lagrangian approaches for calculation of dispersion provide a more natural approach than Eulerian techniques (usually based on eddy-diffusivity concepts which are strictly valid only for the case where the characteristic length scale of the spatial gradients in the mean of a scalar quantity is larger than the integral length scale of the turbulence, implying that these techniques cannot be applied to the simulation of dispersion close to a source). Moreover, LS models can be designed rigorously, based on a small set of principles, to give reliable simulations of dispersion in complex inhomogeneous, non-stationary or non-Gaussian flows where other methods based on eddy diffusivity concepts and/or other semi-empirical approaches (e.g., similarity theory, asymptotic solutions, etc.) are invalid or inadequate. In this regard, the well-mixed condition due to Thomson [1] provides the most rigorously correct theoretical criterion for the formulation of LS models of turbulent dispersion.

In a stochastic or random-walk particle model, dispersion is simulated by following the trajectories of a large number of "marked" fluid particles released from the source into the flow domain, with the mean concentration of the scalar determined from the statistics of the particle displacements. This gives rise to a primary difficulty in the application of LS models; namely, the "correct" and numerically efficient estimation of the mean concentration at a given location and time from the information embodied in the "marked" particle trajectories. A stochastic or random walk model of particle motions is a grid-free method in which particles evolve in time according to stochastic equations of motion. Nevertheless, it is common practice in atmospheric dispersion modelling with LS models to calculate the mean concentration on a fixed grid using a particle-in-cell or box- counting method [2, 3, 4, 5]. In the box-counting method, the mean concentration is proportional to the local particle number density (since the

mass carried by each particle is unchanged along the trajectory of the particle). More specifically, the mean concentration is estimated by multiplying the number of particles in a grid box by the mass carried by each particle (assumed equal), and dividing this quantity by the volume of the grid box. Unfortunately, the box-counting method for mean concentration estimation depends on the choice of the size and position of the grid boxes used for the estimation. To overcome this problem, Yamada et al. [6], Uliasz [7], and de Haan [8] used a density kernel method for mean concentration estimation which fits well into a LS particle model framework since the mean concentration is calculated without reference to a grid. A LS particle model coupled with the use of a density kernel estimator allows the mean concentration to be determined without reference to a grid (and, hence, provides a truly grid-free Lagrangian Monte Carlo method).

In spite of the advantages of the kernel density estimator and the deficiencies in the box-counting method, the latter method is still predominantly used in atmospheric dispersion modelling for the estimation of the mean concentration field. This is because the kernel density estimator for the mean concentration is computationally much more demanding than the simpler box-counting method. More specifically, when implemented in a straightforward manner the computational work of the kernel density estimator scales as $O(N_p N_r)$ (N_p and N_r being the number of particles used in the simulation and the number of receptor points at which the mean concentration is calculated, respectively), whereas in the box-counting method the computational work is of order $O(N_r)$. Since large numbers of particles are typically required to accurately model the mean concentration field, the overall computational cost of using the kernel density estimator in a LS particle model would quickly become prohibitive. In consequence, it would be desirable to develop an algorithm for mean concentration estimation using density kernels for which the computational work scales as $O(N_r)$, enabling a computationally efficient and feasible kernel density method for estimation of concentrations from a stochastic Lagrangian particle model. It is the purpose of this report to develop and implement an algorithm which can calculate the mean concentration from a disordered set of particle positions (obtained from a LS particle model) in $O(N_r)$ computational work as opposed to $O(N_p N_r)$ work. The reduction of the computational complexity using our new algorithm will allow kernel density estimators for mean concentration to be used routinely in LS models for atmospheric dispersion.

Mean Concentration Estimation

The LS modelling of the dispersion of a passive tracer consists of simulating numerically the motion of many particles of tracer ("marked" fluid elements) released from the source distribution, in order to build up a picture of the concentration distribution. More specifically, if $S(\mathbf{x}', t')$ denotes the spatial-temporal density distribution of the source [viz., the mass of source material per unit volume per unit time interval at the source space-time point $(\mathbf{x}', t')$], then the mean concentration $C(\mathbf{x}, t)$ at the receptor space-time point $(\mathbf{x}, t)$ is determined from [9]

(1)
$$C(\mathbf{x}, t) = \int_V \int_0^t p_L(\mathbf{x}, t | \mathbf{x}', t') S(\mathbf{x}', t') \, dt' \, d\mathbf{x}',$$

where $p_L(\mathbf{x},t|\mathbf{x}',t') \equiv \langle \delta(\mathbf{x} - \mathbf{X}(t;\mathbf{x}',t')) \rangle$ is the conditional probability density function (PDF) that a marked particle released at the source space-time point $(\mathbf{x}',t')$ will be found at receptor (observation) space-time point $(\mathbf{x},t)$. Here, $\delta(\cdot)$ is the Dirac delta function, $\langle \cdot \rangle$ is used to denote the averaging over the samples of the turbulent velocity field, and $\mathbf{X}(t;\mathbf{x}',t')$ denotes the Lagrangian spatial coordinates defining the trajectory in physical space of a particle "marked" at the source point $(\mathbf{x}',t')$. In Eq. (1), the spatial integration is over the domain (or, volume) V that contains the emission-source distribution $S(\mathbf{x}',t')$.

In accordance to Eq. (1), to estimate the mean concentration at receptor point $\mathbf{x}$ and time t requires the representation of the integral here as an expectation of a random estimator defined on Lagrangian trajectories. Since the exact form of $p_L(\mathbf{x},t|\mathbf{x}',t')$ is unknown, we need to approximate this particle displacement PDF from the solutions of stochastic evolution equations describing trajectories of fluid particles in a turbulent flow. The evolution equations for the position $\mathbf{X}(t)$ and velocity $\mathbf{U}(t)$ of a tagged fluid particle in a Lagrangian Stochastic model is typically based on a continuous Markovian evolution in position-velocity (phase) space specified by a stochastic differential equation of the form [1]

$$(2) \qquad dU_i = a_i(\mathbf{U},\mathbf{X},t)dt + \left(C_0\varepsilon\right)^{1/2}d\xi_i(t)$$

and

$$(3) \qquad dX_i = U_i dt,$$

where dU_i and dX_i ($i = 1,2,3$) denote the change (or, increment) in velocity and position over the time interval dt, respectively. More specifically, $dU_i = U_i(t+dt) - U_i(t)$ is the infinitesimal increment of the velocity U_i following the marked fluid particle. Appearing in the equations is the drift coefficient vector $a_i(\mathbf{U},\mathbf{X},t)$, the universal constant C_0 associated with the Lagrangian velocity structure function, the mean rate of dissipation of turbulence kinetic energy ε, and the isotropic vector incremental Weiner process $d\xi_i(t)$. The value $C_0 = 3.5$ is recommended for the LS model used here [2, 10]. The increments $d\xi_i$ have a joint-normal distribution with zero means and an isotropic covariance matrix,

$$(4) \qquad \langle d\xi_i(t) \rangle = 0, \qquad \langle d\xi_i(t)d\xi_j(t+\tau) \rangle = dt\delta_{ij}\delta(\tau),$$

where δ_{ij} denotes the Kronecker delta function. Finally, it should be emphasized that in a "correctly" formulated LS model, the drift coefficient vector should be chosen to satisfy the well-mixed condition [1] which guarantees that the model will maintain the correct phase-space distribution of particles in the sense that once the particles of tracer are well-mixed in the flow, they will remain so.

The conventional method for evaluating the mean concentration of Eq. (1) is to apply the box-counting method and count the number of particles in a discrete grid cell centered at $\mathbf{x}$ at time t. This process is identical to the calculation of a three-dimensional histogram. In particular, the estimate $\hat{C}$ for the mean concentration at a given location $\mathbf{x}$ and time t is given by

$$(5) \qquad \hat{C}(\mathbf{x},t) = \frac{N_{pl}(t)}{\Delta_l} \times \Delta m.$$

where $N_{pl}(t)$ is the number of "marked" fluid particles at time t in the spatial cell l (that contains the receptor position $\mathbf{x}$), Δ_l is the volume of cell l, and Δm is the mass of tracer carried

by each fluid particle. Each fluid particle in the ensemble of N_p particles is assumed to carry a constant mass Δm. Notice that, for a given spatial cell l, the sample size is usually relatively small (i.e., $N_{pl} \ll N_p$), and hence the statistical error in the concentration estimate $\hat{C}(\mathbf{x},t)$ will generally be relatively large with the error scaling as $N_p^{-1/2}$. In consequence, box-counting estimates for the mean concentration will therefore generally require large sample sizes (i.e., N_p large) if the statistical error is to remain small. Alternatively, it is possible to reduce the statistical error by increasing the size (volume) of the spatial cell, but this may result in the structural details in the mean concentration being oversmoothed (leading to a larger bias in the mean concentration estimate). In summary, many particles per spatial cell are required to keep the bias and statistical error in the mean concentration estimate small, but this requirement makes trajectory simulations computationally expensive and hence prohibitive.

For these reasons, Uliasz [7] and de Haan [8] proposed another method for the estimation of the mean concentration based on density kernels. The foundation of the kernel estimate of a field variable at a point is interpolation theory generalized to the case of interpolation from the disordered positions of an ensemble of marked particles, where the underlying turbulent velocity field "deals" out the disordered positions. Within the volume of the physical space V, the mean concentration of the dispersing scalar is represented by N_p particles, each representing a mass of tracer Δm with the n-th particle—the numbering being arbitrary—having a position $\mathbf{X}^{(n)}(t)$ and velocity $\mathbf{U}^{(n)}(t)$ at time t. Then the mass density function (concentration) $C^*(\mathbf{x},t)$ (which can be interpreted to be the discrete representation of the mean concentration distribution $C(\mathbf{x},t)$ of a dispersing scalar and, as such can be considered to be an estimate of this distribution) is defined by

$$(6) \qquad C^*(\mathbf{x},t) \equiv \Delta m \sum_{n=1}^{N_p} \delta\big(\mathbf{x} - \mathbf{X}^{(n)}(t)\big),$$

where

$$(7) \qquad \delta\big(\mathbf{x} - \mathbf{X}^{(n)}(t)\big) \equiv \delta\big(x_1 - X_1^{(n)}(t)\big)\delta\big(x_2 - X_2^{(n)}(t)\big)\delta\big(x_3 - X_3^{(n)}(t)\big)$$

is the 3-dimensional Dirac delta function. For simplicity (but without any loss in generality), assume that the tracer is released from an instantaneous point source located at the origin $\mathbf{x} = \mathbf{0}$ with the release occurring at time $t = 0$. Denoting by Q the total mass of tracer released ($Q = N_p\Delta m$), the (mathematical) expectation of Eq. (6) is

$$(8) \qquad \langle C^*(\mathbf{x},t)\rangle = Q\langle\delta\big(\mathbf{x} - \mathbf{X}^*(t)\big)\rangle,$$

where $*$ refers to any particle n ($1 \leq n \leq N_p$). Comparing Eq. (8) with Eq. (1) for the case of an instantaneous point source release at $\mathbf{x} = \mathbf{0}$ and $t = 0$ ($S(\mathbf{x}',t') = Q\delta(\mathbf{x}')\delta(t')$), it is straightforward to show that the mass density function in Eq. (6) is a consistent (unbiased) representation of the mean concentration in the sense that

$$(9) \qquad \langle C^*(\mathbf{x},t)\rangle = C(\mathbf{x},t).$$

The use of an interpolation function that provides the kernel estimate of the mean concentration at a space-time point involves replacing the 3-dimensional Dirac delta function in Eq. (6) by a weighting function (or, "kernel") that defines how much of the mass carried by a "marked" fluid particle contributes to the mean concentration field at the point $\mathbf{x}$ at time t. The kernel estimate of $C(\mathbf{x}.t)$ is then given by

$$(10) \qquad \hat{C}(\mathbf{x},t) \equiv \Delta m \sum_{n=1}^{N_p} \frac{1}{h^3} K\left(\mathbf{x} - \mathbf{X}^{(n)}(t); h\right),$$

where h is the width of the (interpolating) kernel function $K(\mathbf{r}; h)$. The kernel function is a non-negative valued function $[K(\mathbf{r}; h) \geq 0]$ that satisfies the property

$$(11) \qquad \int_{\text{a.s.}} K(\mathbf{r}; h)\, d\mathbf{r} = 1,$$

where the integral is taken over all space (a.s.).

The kernel method does not require the physical space to be partitioned into grid cells and produces a smooth mean concentration field with a much smaller number of particles than is required for the box-counting method. The width (or, smoothing length) h determines the degree of smoothing of the mean concentration field. The smaller values of h give a more local estimate for the mean concentration, but result in fewer particles giving significant contributions at a given (receptor) point $\mathbf{x}$ and, hence, more statistical error in the estimate. Taking h larger allows more "marked" fluid particles to contribute to the mean concentration estimate at a receptor location $\mathbf{x}$ reducing the statistical error, but taking h too large will smooth out detailed features from the mean concentration distribution increasing the bias in the estimate. In consequence, the choice of the bandwidth h is a compromise between smoothing enough and not smoothing too much to smear out the real features in the mean concentration distribution. Mathematically, there is a compromise between the bias and variance of $\hat{C}(\mathbf{x},t)$, which increases and decreases, respectively, as h is increased. Theory for density kernel estimators [11] suggests that the kernel bandwidth should be chosen proportional to $N_p^{-1/5}$, but the constant of proportionality depends on the unknown mean concentration distribution.

Following de Hann [8], we consider two types of kernel functions K with finite (or, compact) support. The latter implies that only a subset of the "marked" fluid particles contributes to each kernel estimate of the mean concentration. In particular, we use kernel functions with the general form

$$(12) \qquad K(\mathbf{r}, h) = \begin{cases} C_{d,a}\left(1 - s^2\right)^a; & s^2 < 1, \\ 0; & s^2 \geq 1, \end{cases}$$

where $s \equiv (\mathbf{r} \cdot \mathbf{r})^{1/2}/h$, a is an exponent, d is the number of dimensions, and $C_{d,a}$ is the normalization constant. In this report, we consider kernel functions with $a = 1$ (parabolic kernel, which is referred to also as Epanechnikov kernel [12]) and $a = 4$ (quadweight kernel). For these two kernels, the normalization constant $C_{d,a}$ [obtained from the normalization constraint Eq. (11)] takes the values $\frac{2}{\pi}$ or $\frac{15}{8\pi}$ for $a = 1$ and $\frac{5}{\pi}$ or $\frac{3465}{512\pi}$ for $a = 4$ in two- and

three-dimensions ($d = 2$ and 3), respectively. One of the key advantages of using Eq. (12), as opposed to, say a Gaussian kernel function ($K(\mathbf{r}; h) = \exp(-s^2/2)/(2\pi h)^{d/2}$), is that it has a finite support so that "marked" fluid particle contributions to mean concentration at a given (receptor) point $\mathbf{x}$ are exactly zero for particles at distances $r \equiv (\mathbf{r} \cdot \mathbf{r})^{1/2} > h$ (or, equivalently, $s \equiv r/h > 1$). Figure 1 shows the profiles of non-normalized parabolic and quadweight kernels (i.e., profiles of $K(\mathbf{r}, h)/C_{d,a} \equiv \mathrm{K}(s)$).

A Fast Algorithm for Kernel Density Estimation

Algorithm description

The näive implementation of the kernel estimate [cf. Eq. (10)] for the mean concentration field is simple. Since the kernel functions considered here have a compact support [refer to Eq. (12)], only a subset of the fluid particles will contribute to the kernel estimate at a given (fixed) receptor location $\mathbf{x}$, corresponding as such to the nearest neighboring particles for the receptor position. The näive evaluation of Eq. (10) then consists of an all-pair search approach whereby the distance $r^{(n)} \equiv \left((\mathbf{x} - \mathbf{X}^{(n)}(t)) \cdot (\mathbf{x} - \mathbf{X}^{(n)}(t)) \right)^{1/2}$ ($n = 1, 2, \ldots, N_p$) of each tagged fluid particle from the receptor location $\mathbf{x}$ is calculated, and if this distance is smaller than h (bandwidth of the kernel function centered about the particle position, which also defines the support domain of the function), the mass carried by the particle contributes to the mean concentration at receptor point $\mathbf{x}$. The all-pair search approach needs to be carried out for each receptor location $\mathbf{x}_r$ ($r = 1, 2, \ldots, N_r$, where N_r is the number of receptor locations where the mean concentration is calculated), and at each of these locations the searching is performed for all N_p particles at the given time t. It is clear that the complexity of the all-pair search algorithm for the computation of the kernel estimate of the mean concentration field is of order $O(N_r N_p) \sim O(N^2)$ (if $N_r \approx N_p = N$). Furthermore, this näive algorithm needs to be applied at all time steps (since the tagged fluid particle positions are evolving in time) so that the computational cost of the conventional (albeit straightforward) implementation of the kernel estimate of the mean concentration field is computationally prohibitive (and, typically, too expensive to be applied in practical problems where the number of particles is large).

The straightforward implementation of the kernel estimate for the mean concentration corresponds to a huge waste of computational time since only a small fraction of the total N_p fluid particles will contribute to the concentration at a given (fixed) receptor location. Our approach for improving the computational efficiency of the kernel estimator (grid-free method) is to maintain a neighbor list for each fluid particle. The methodology used to keep track of nearest neighbor fluid particles relative to a given receptor point is to divide the physical volume V (flow domain) into cells or elements (viz., a temporary mesh is overlaid on the flow domain to provide a tesselation of the domain). Each cell in the tesselation of the 3-dimensional domain is a cube (or, a square for a 2-dimensional domain) with side length equal to h (smoothing length of the kernel function), the reason being that the kernel functions centered on the fluid particles only contribute to the concentration at a receptor location if the distance r between the particle position and the receptor location is less than h (i.e., $r < h$). In consequence, with cells of side length h, we only need to check the current cell which contains the receptor location and nearest neighboring cells in order to find all possible fluid particles

that can contribute to the mean concentration at the receptor location.

To facilitate using cells as a book-keeping device to keep track of fluid particle positions, each cell in the superimposed structured grid is associated with a linked list. Each "marked" fluid particle in the flow domain is assigned to the cell it resides in and identified through the linked list associated with that cell. The cells in the tesselation of the physical domain can be identified by the index set (i, j) or (i, j, k), respectively, for a 2-dimensional or 3-dimensional domain. If (i, j, k) is the index set for cell P in a 3-dimensional physical region, then the cells immediately to the right and left of P are labelled $(i + 1, j, k)$ and $(i - 1, j, k)$, the cells immediately in front and back are $(i, j + 1, k)$ and $(i, j - 1, k)$, and the cells directly above and below are $(i, j, k + 1)$ and $(i, j, k - 1)$. For a given receptor location $\mathbf{x}$ contained in cell P with label (i, j, k), all fluid particles which can contribute to the concentration at $\mathbf{x}$ can only reside in cell P and/or in the one of its immediately adjoining (nearest neighbor) cells with labels $(i \pm \alpha, j \pm \beta, k \pm \gamma)$, where α, β, $\gamma \in \{0, 1\}$ excluding the choice $\alpha = \beta = \gamma = 0$ (which corresponds to cell P itself). Hence, for the computational cell (box) P labelled by the index (i, j, k), the list of all nearest neighbor cells to P, denoted by near(P), is the set of all cells labelled by the index set $\{(i \pm \alpha, j \pm \beta, k \pm \gamma) | \alpha, \beta, \gamma \in \{0, 1\}, \alpha \neq \beta \neq \gamma = 0\}$. Therefore, the search for fluid particles that can contribute to the concentration at $\mathbf{x}$ is confined to only 9 or 27 cells in 2- or 3-dimensional space, respectively. The linked list associated with each cell allows the fluid particles assigned to each cell to be chained together for easy access.

The superimposed grid allows the "binning" (or sorting) of the fluid particles into N_c cells, with the average number of particles per cell being about N_p/N_c (assuming the particles are uniformly distributed in the physical domain). If the average number of fluid particles per cell is small ($N_p \gg N_p/N_c$, the latter of which is of order $O(1)$), the computational complexity of the linked list algorithm for the kernel estimate of the mean concentration field is of $O(N_r)$. This computational effort is optimal since the program must at least loop through all N_r receptor locations in order to compute the mean concentration field at a particular time t (viz., traversing a "data" set of N_r receptor locations has a minimal computational complexity of order $O(N_r)$).

Programming considerations

A computer program that implements the linked list (fast) algorithm for the kernel estimation of the mean concentration field has been written. The program contains less than several hundred lines of Fortran 90/95 code. In this subsection, we summarize some of the more technical details of our implementation of the fast kernel density estimator.

The beginning of the code shown below contains definitions of the components for one derived data type: namely, cells. This derived data type cells contains an integer variable counter to hold the fluid particle number index, four real variables mass, x, y, and z to hold the fluid particle mass and the coordinates (x, y, z) of the fluid particle position in 3-dimensional space, and a pointer variable NextParticle to a derived data type cells variable as its last component. This pointer points to the next item (particle) in the linked list of particle properties contained in a given computational cell. A linked list of particles associated with a

given cell is simply a series of variables in the derived data type `cells` with the pointer from each derived data type variable pointing to the next variable in the list. The pointer `NextParticle` of the last element of the linked list is set to `NULL`, a convenient sentinel marking the end of the list.

Each cell of the tesselation of the physical region has a linked list associated with it, the latter used simply as a convenient book-keeping device to keep track of the particles inside that cell at a given time t. This feature is very useful because it permits us to construct various types of dynamic data structures linked together by successive pointers during the execution of a program. We use a linked list to hold the fluid particle properties in a cell because the number of fluid particles in a given cell can vary with time as the position and velocity of the particles evolve according to the stochastic differential equations exhibited in Eqs. (2) and (3), and the linked list permits us to add (remove) elements to (from) the list one at a time without knowing in advance how many elements will ultimately be in the list. Since there are mutiple cells in the flow domain, another derived data type `cell_pointer` is defined and used in the declaration of 3-dimensional arrays of pointers to derived data type `cells`. Here, `cell_pointer` is defined to be a pointer to the derived data type `cells`. Finally, two 3-dimensional arrays of pointers (`head` and `tail`) to variables of derived data type `cells` are also defined to point to the first and last variables in the linked list for each cell. To access an element in the linked list, we start at the head of the list and use the pointers `NextParticle` of successive elements to move from element to element until the desired element is reached. With the singly-linked lists used here, the list can be traversed in only one direction (from head to tail) because each element contains no link to its predecessors.

The following is a formal definition of some derived data types required by the algorithm.

```fortran
TYPE :: cells
    INTEGER :: counter
    REAL :: mass
    REAL :: x
    REAL :: y
    REAL :: z
    TYPE (cells), POINTER :: NextParticle
END TYPE

TYPE cell_pointer
    TYPE(cells), POINTER :: pptr
END TYPE cell_pointer

TYPE(cell_pointer), ALLOCATABLE, DIMENSION(:,:,:) :: head
TYPE(cell_pointer), ALLOCATABLE, DIMENSION(:,:,:) :: tail

TYPE (cells) :: temp                ! Temporary variable
```

Having defined some useful data structures, the fast algorithm for kernel estimation of the

mean concentration field consists of the following main steps:

1. Tesselate the flow domain into cells (cubes) with side length h (bandwidth of the kernel function).

2. Open the input file containing the "marked" fluid particle properties (index, mass, location) at a given time t.

3. Read in the properties (index, mass, location) of each fluid particle, determine the cell the particle resides in, and insert the relevant particle information into the linked list for that cell.

4. Define the set of receptor locations at which to calculate the mean concentration.

5. For each receptor location, determine the cell P that it resides in, extract the fluid particle properties from the linked list for this cell, and calculate the contribution of each of these fluid particles to the kernel density estimator of the mean concentration.

6. Identify the 26 nearest neighbor cells ($near(P)$) to the cell P containing the receptor location, extract the fluid particle properties from the linked lists for each of these adjoining cells, and for each particle calculate the Euclidean distance r between the particle and receptor location. For each particle whose distance r from the receptor position is less than h, calculate the contribution of this particle to the kernel density estimator of the mean concentration.

7. Output the mean concentration field at the given time t.

Following is a formal description of the algorithm.

Algorithm

```
Input functional form of kernel function
Input bandwidth h of kernel function
Subdivide the flow domain into contiguous cells of side length h

Open file containing fluid particle property information

WHILE
    Read fluid particle property information into temp
    IF read not successful EXIT
    Determine index of cell where current particle resides
    Get linked list for current cell
      ! determine which cell (ip,jp,kp) the particle is located in

        ip = INT((temp%x - Xmin)/bandwidth) + 1
```

```
                jp = INT((temp%y - Ymin)/bandwidth) + 1
                kp = INT((temp%z - Zmin)/bandwidth) + 1

        IF head(ip,jp,kp) (of linked list) is not associated THEN

          ! The list is empty

                ALLOCATE head(ip,jp,kp)

          ! Tail points to first value

                tail(ip,jp,kp) => head(ip,jp,kp)

          ! Nullify pointer NextParticle within 1st value
          ! since there is nothing to point to yet

                NULLIFY tail(ip,jp,kp)%NextParticle

          ! Store particle information (index,mass,location)
          ! in linked list

                tail(ip,jp,kp) <-- temp

        ELSE
          ! The list already has particle information inserted

                ALLOCATE tail(ip,jp,kp)%NextParticle

          ! Tail now points to new last value

                tail(ip,jp,kp) => tail(ip,jp,kp)%NextParticle

          ! Nullify pointer within new last value

                NULLIFY tail(ip,jp,kp)%Nextparticle

          ! Store new particle information

                tail(ip,jp,kp) <-- temp

        END of IF

END of WHILE

Define receptor locations and calculate index (ic,jc,kc)
```

This page intentionally left blank.

of cell where each receptor resides

```
DO for each receptor location with index (ic,jc,kc)

    Get linked list for receptor cell

        ptr => head(ic,jc,kc)     ! go back to the head of the list
        WHILE ptr is associated

          Get the particle coordinates
          Calculate contribution to concentration at receptor
          ptr => ptr%Nextpointer

        END of WHILE

    Get linked lists for all the nearest neighbor cells to
    receptor cell P with index (ic,jc,kc)

    DO for each linked list in near(P)
        ptr => head(near(P))     ! go back to the head of the list
        WHILE ptr is associated
            Get the particle coordinates
            Calculate Euclidean distance from particle to receptor
            If distance less than h, calculate contribution
            to concentration at receptor
            ptr => ptr%Nextpointer
        END of WHILE
    END of DO

END of DO   ! next receptor location

WRITE out mean concentration field at all receptor points
```

A Fortran 90/95 computer program that implements the fast kernel estimator for the mean concentration is available from the authors upon request. The computer program has been implemented using both 2-dimensional and 3-dimensional kernels.

Simple Lagrangian Stochastic Particle Model

As a test of our new method for the fast kernel estimation of the mean concentration, we have written a simple trajectory-simulation model for turbulent dispersion from an instantaneous and continuous point source in a neutrally stratified inhomogeneous shear flow (constant stress

layer in the atmospheric boundary layer). The trajectory-simulation approach to modelling dispersion of a passive tracer consists of simulating the trajectories of many particles of tracer to build up a picture of the mean concentration distribution. In this section, we use Cartesian tensor notation, but for convenience ocassionally we also use Cartesian coordinates (x,y,z) and (u,v,w) to denote the streamwise, cross-stream, and vertical coordinates and velocities, respectively.

Particle model description

The dispersion of a passive scalar is described by the stochastic differential equation (generalized Langevin equation) exhibited in Eqs. (2) and (3). The drift term (or, conditional particle acceleration) a_i in Eq. (2) is constrained for a particular flow by specifying the the Eulerian velocity statistics for that flow, and requiring that the Eulerian statistics determined from the LS model [Eqs. (2) and (3)] reproduce the specfied statistics. A sufficient condition for this to be achieved is embodied in the well-mixed criterion (Thomson [1]) which requires the Eulerian velocity PDF $P_E(\mathbf{u};\mathbf{x},t)$ satisfy the Fokker-Planck equation for the joint PDF of velocity and position $P(\mathbf{u},\mathbf{x},t)$ associated with the stochastic differential equation system specfied by Eqs. (2) and (3). Application of the well-mixed criterion constrains a_i as

$$(13) \qquad a_i P_E = \frac{1}{2} C_0 \varepsilon \frac{\partial P_E}{\partial u_i} + \phi_i,$$

where

$$(14) \qquad \frac{\partial \phi_i}{\partial u_i} = -\frac{\partial P_E}{\partial t} - \frac{\partial u_i P_E}{\partial x_i},$$

and $\phi_i \to 0$ as $|\mathbf{u}| \to \infty$ ($|\cdot|$ denotes the Euclidean norm). Note that Eq. (14) does not determine ϕ_i uniquely since an arbitrary solenoidal vector function in velocity space (i.e., function whose divergence with respect to u_i vanishes) can be added to ϕ_i and still satisfy Eq. (14). Hence, the well-mixed criterion allows one to determine 'a' drift function rather than 'the' drift function since one can add any solenoidal vector function ψ_i ($\partial \psi_i / \partial u_i = 0$) to ϕ_i without changing $\partial \phi_i / \partial u_i$ and the interpretation of this "gauge" freedom in terms of the statistical properties of the flow is awkward.

We consider the case of Gaussian stationary, inhomogeneous turbulence where P_E is given by

$$(15) \qquad P_E(\mathbf{u};\mathbf{x},t) = \frac{\gamma}{(2\pi)^{3/2}} \exp\left(-\frac{1}{2}(u_i - \bar{U}_i)\Gamma_{ij}(u_j - \bar{U}_j) \right),$$

where $\Gamma_{ij} \equiv (V^{-1})_{ij}$ is the inverse of the Reynolds stress tensor V_{ij} (i.e., $V_{ij} \equiv \langle u_i' u_j' \rangle$ is the covariance between the i-th and j-th components of the Eulerian velocity where a primed quantity refers to the deviation from the mean value for that quantity), γ is the determinant of Γ_{ij}, and $\bar{U}_i$ is the mean velocity. Summation over repeated tensor indices $i,j,k,\ldots$ is implied (Einstein summation convention). In this case, Thomson [1] provided the following form for a conditional acceleration vector ϕ_i (with $u_i' \equiv u_i - \bar{U}_i$ being the velocity fluctuation) that is consistent with the well-mixed criterion:

$$(16) \qquad \frac{\phi_i}{P_E} = \bar{U}_l \frac{\partial \bar{U}_i}{\partial x_l} + \frac{\partial \bar{U}_i}{\partial x_l} u_l' + \frac{1}{2} \frac{\partial V_{il}}{\partial x_l} + \frac{1}{2} \bar{U}_m \frac{\partial V_{il}}{\partial x_m} \Gamma_{lj} u_j' + \frac{1}{2} \frac{\partial V_{il}}{\partial x_k} \Gamma_{lj} u_j' u_k'.$$

The expanded form of Eq. (16) is complicated. In this report, we consider a simple neutrally-stratified flow with mean velocity vector $\bar{\mathbf{U}} = (\bar{U}, 0, 0)$ and turbulence velocity statistics varying only in the z-direction (vertical direction). Furthermore, we make the assumption that we can ignore the covariance between the different velocity components (i.e., $\langle u'_i u'_j \rangle = 0$ for $i \neq j$). In this case, Thomson's solution for the drift coefficient reduces to with $i = 1, 2, 3 \equiv u, v, w$:

$$a_u = -\frac{C_0 \varepsilon}{2\sigma_u^2} u' + \frac{\partial \bar{U}}{\partial z} w' + \frac{1}{2\sigma_u^2} \frac{\partial \sigma_u^2}{\partial z} u' w',$$

$$a_v = -\frac{C_0 \varepsilon}{2\sigma_v^2} v' + \frac{1}{2\sigma_v^2} \frac{\partial \sigma_v^2}{\partial z} v' w',$$

(17)
$$a_w = -\frac{C_0 \varepsilon}{2\sigma_w^2} w' + \frac{1}{2} \frac{\partial \sigma_w^2}{\partial z} \left(\frac{w'^2}{\sigma_w^2} + 1 \right),$$

where σ_u^2, σ_v^2, and σ_w^2 are the Eulerian velocity variances in x-, y-, and z-directions, respectively.

The meteorological inputs of the model are the vertical profiles of the mean wind speed $\bar{U}$, of the Eulerian velocity variances σ_u^2, σ_v^2, and σ_w^2, and the turbulence kinetic energy (TKE) mean dissipation rate ε. In this report we consider only a neutrally stratified shear flow (i.e., Obukhov length $L \to \infty$), and we assume that the Eulerian velocity variances in the x- and y-directions (horizontal velocity variances) are equal so $\sigma_u^2 = \sigma_v^2$. The average wind speed in the surface layer above the ground is described analytically by the semi-logarithmic law-of-the wall:

(18)
$$U(z) = \frac{u_*}{\kappa} \log \left(\frac{z}{z_0} \right),$$

where u_* is the friction velocity, $\kappa \approx 0.4$ is von Karman's constant, and z_0 is the surface roughness length. Our parameterizations for the turbulence velocity flow statistics follow those suggested by Rodean [13] for a neutrally-stratified boundary-layer flow:

(19)
$$\sigma_u^2 = \sigma_v^2 = u_*^2 \left[4.5 \left(1 - \frac{z}{H} \right)^{1.5} \right],$$

(20)
$$\sigma_w^2 = u_*^2 \left[2.0 \left(1 - \frac{z}{H} \right)^{1.5} \right],$$

(21)
$$\varepsilon = \frac{u_*^3}{\kappa z} \left(1 - 0.85 \frac{z}{H} \right)^{1.5},$$

where H is the depth of the (atmospheric) boundary layer.

The stochastic differential equations (2) and (3) with the drift coefficients (conditional acceleration) determined from Eq. (17) are integrated forward in time using a simple first-order Euler scheme. With this scheme, the update from time t_n to time t_{n+1} (with time step $\Delta t = t_{n+1} - t_n$) for advancing the fluid particle position and velocity has the form

(22)
$$\mathbf{X}^{n+1} = \mathbf{X}^n + \mathbf{U}^n \Delta t,$$

(23)
$$\mathbf{U}^{n+1} = \mathbf{U}^n + \mathbf{a}^n \Delta t + \left(C_0 \varepsilon^n \right)^{1/2} \Delta \mathbf{W}^n,$$

where ΔW_i is a Gaussian pseudo-random number with mean $\langle \Delta W_i \rangle = 0$ and covariance $\langle \Delta W_i \Delta W_j \rangle = \Delta t \delta_{ij}$. Superscripts n and $n+1$ in Eq. (23) denote the values at time t_n and $t_{n+1} = t_n + \Delta t$, respectively, and $\mathbf{a}^n$ is the drift coefficient vector evaluated at time t_n, i.e. $\mathbf{a}^n = \mathbf{a}(\mathbf{U}^n, \mathbf{X}^n, t_n)$. Note that the coefficients in Eq. (23) are evaluated at $\mathbf{X}^n$ and $\mathbf{U}^n$ and, hence, the Euler approximation is explicit. The time step Δt is chosen to be a small fraction of the Lagrangian time scale $\tau_L \equiv 2\sigma_w^2/(C_0\varepsilon)$ (more specifically, a time step $\Delta t = 0.01\tau_L$ was found to give both an accurate and stable solution to the stochastic system for fluid particle evolution). The value of the universal constant C_0 possesses considerable uncertainty. The existing literature seems to suggest that it depends on the structure of the turbulent flow being considered. As regards to the choice of C_0 for atmospheric dispersion applications, it appears that a choice $C_0 \approx 3.5$ provides good agreement with data [2, 10] for the one-dimensional LS model (model that neglects the shear stress of the background flow).

Simulation Results

Continuous point source

In the first set of numerical experiments, we test the code with experimental data and compare the estimate for the mean concentration obtained using the box counting and the fast kernel density estimator methods with measured data. To this purpose, concentration profiles measured in the atmospheric surface layer under near neutral conditions during the Project Prairie Grass experiment will be used as a reference for comparison.

Project Prairie Grass represents a benchmark field experiment for turbulent dispersion in the atmospheric surface layer. The Project Prairie Grass (PPG) experiment is fully discussed in the reports by Barad [14] and Haugen [15]. A continuous point source of sulfur dioxide (SO_2) at a height of 0.46 m above an extensive flat plain was used in this experiment. In each run, the time-averaged (mean) concentration was measured over a sampling period of 10 min using a large number of downwind detectors placed on circular arcs centered at the source. All data considered herein are from the arc of detectors with a radius of 100 m as this was the only arc at which the vertical concentration profile was measured. Equivalent two-dimensional concentration profiles $C^y(x,z)$ with $x = 100$ m were derived by performing a crosswind integration on the concentration; that is, the crosswind integrated concentration (CWIC) was obtained from

$$(24) \qquad\qquad C^y(x,z) = \int C(x,y,z)\,dy.$$

Sulfur dioxide is almost a perfect conservative tracer with the uptake by the (grass) surface being small (i.e., the surface is almost a perfect reflector). We therefore treated the surface as a perfect reflector at the roughness length z_0 when calculating fluid element trajectories. In the PPG experiment, the roughness length of the surface was estimated to be about $z_0 \approx 0.006$ m. In application of the kernel estimator for the mean concentration, the concentration at a given point near the surface must account explicitly for the reflection of particles from the ground

surface, with the result that the kernel estimate for $C(\mathbf{x},t)$ given by Eq. (10) is replaced by

$$(25) \qquad \hat{C}(\mathbf{x},t) \equiv \Delta m \sum_{n=1}^{N_p} \frac{1}{h^3} \left[K\left(\mathbf{x} - \mathbf{X}^{(n)}(t); h\right) + K\left(\mathbf{x} - \mathbf{X}^{(n)*}(t); h\right) \right],$$

where $\mathbf{X}^{(n)*} = (X_1^{(n)}, Y_1^{(n)}, -Z_1^{(n)})$ is the location of a virtual fluid particle that is associated with the real fluid particle at the location $\mathbf{X}^{(n)} = (X_1^{(n)}, Y_1^{(n)}, Z_1^{(n)})$ (viz., the virtual fluid particle is the mirror image of the real fluid particle with respect to the ground surface at $z = 0$). More specifically, for particles near the ground surface (at a vertical distance less than the kernel bandwidth h) at $\mathbf{X}^{(n)} = (X_1^{(n)}, Y_1^{(n)}, Z_1^{(n)})$ (with $Z_1^{(n)} < h$), an equivalent virtual (or, ghost) particle is also introduced at $\mathbf{X}^{(n)*} = (X_1^{(n)}, Y_1^{(n)}, -Z_1^{(n)})$ to provide a symmetrical surface boundary condition at $z = 0$. It is noted that for the continuous release considered here, the position of the 'marked' fluid particles are written to the data file every δT s for the entire simulation time, and this information is used to constuct a kernel estimate for the mean concentration exhibited in Eq. (25). In particular, each 'marked' fluid particle carries the source mass flow rate Q/N_p (Q is the mass flow rate of the source) so the mass carried by each particle over the interval δT is $\delta m = Q\delta T/N_p$). For box-counting, the steady-state mean concentration in a small cell ($\Delta x \times \Delta y \times \Delta z$) enclosing the receptor point (x,y,z) is estimated as follows:

$$(26) \qquad \hat{C}(x,y,z) = \frac{Q}{N_p \Delta x \Delta y \Delta z} \sum_{n=1}^{N_p} T_n,$$

where Q is the source strength (kg s^{-1}) and T_n is the total time the n-th particle remains in the cell (residence time) and is calculated by accumulating the time steps dt whenever the n-th particle resides in the cell.

Figure 2 shows the observed and predicted vertical profile of normalized cross wind-integrated concentration $u_* C^y / Q$ at a downwind fetch $x = 100$ m due to a continuous source of strength Q (g(SO$_2$)s^{-1}) releasing tracer into a near-neutral surface layer. With this normalization, the normalized crosswind integrated concentration is independent of u_* (for a fixed z_0 and experimental geometry). The observations here represent the average from the nine PPG runs that were the nearest-neutral in atmospheric stratification (namely, the average from PPG runs 49, 62, 26, 61, 30, 20, 33, 45, and 57 with Obukhov lengths L in the range from -36 to -240 m). The predicted results for mean concentration were estimated from 5,000 independent particle trajectories. Based on only 5,000 independent trajectories, the fast kernel density estimator (using the quadweight kernel function) for the mean concentration yielded a good conformance with the PPG observations. The mean concentration estimate based on the box-counting method (using bins $\Delta x = 2$ m and $\Delta z = 0.1$ m) also gave results that closely resembled the PPG data, although with only 5,000 independent trajectories the statistical error (or, variability) in the estimate is quite significant. In the box-counting method, the root-mean-square statistical error scales as $N_p^{-1/2}$.

Figure 3 is a plot showing the mean concentration estimate based on box-counting using 50,000 independent particle trajectories. The statistical error in this box-counting estimate for the mean concentration is reduced considerably in comparison with that obtained using only

5,000 particle trajectories. Furthermore, note that the kernel estimate for the mean concentration based on 5,000 independent particle trajectories gives as good a prediction for the concentration as that based on box counting using 50,000 particle trajectories.

Instantaneous point source

In this subsection, we consider the tracer dispersion resulting from an instantaneous point source released in the neutrally-stratified, horizontally homogeneous, wall-shear layer (constant stress layer). To simulate the mean concentration field resulting from an instantaneous point source, thousands of particles are typically released at the source location with their initial velocities equal to the Eulerian turbulent velocities at this location, and their random trajectories calculated using Eqs. (2) and (3), with the drift term a_i ($i = 1, 2, 3 \equiv u, v, w$) specified by Eq. (17). In the following simulations, an instantaneous point source located at $x = y = 0$ and $z = 1$ m releases $Q = 1$ kg of tracer into the neutrally-stratified constant stress layer with a roughness length $z_0 = 0.006$ m and friction velocity $u_* = 0.1 \ \mathrm{ms}^{-1}$.

Figures 4 and 5 show comparisons of the mean concentration estimate along the mean cloud centerline at $y = 0$ and at a height $z = 1$ m above ground level obtained using the direct (näive) implementation of the density kernel estimator and the fast kernel estimator algorithm described herein. The concentration estimates were based on 50,000 independent fluid particle trajectories, and were obtained at six different times after the instantaneous point source release (at $t = 0$). The density kernels used here were the parabolic (Epanechnikov) (Figure 4) and quadweight (Figure 5) kernel functions applied with a bandwidth $h = 2$ m. Note that the direct and fast algorithms for kernel density estimation gave identical results for the mean concentration (as they should). However, for the calculation of the mean concentration field at 51^3 receptor locations for 6 different travel times, the fast algorithm and direct method for density kernel estimation required, respectively, 14 s and 2800 s to complete. The calculations cited here were carred out on a PC with a Xeon processor (3.2 GHz clock speed) running Red Hat Linux 9 Operating System. For this example, the fast algorithm is roughly 200 times faster than the direct method for density kernel estimation, and this speedup factor is expected to increase for applications where the number of receptor locations and/or independent fluid particle trajectories are greater than what were used for this simple example.

It was shown in the previous subsection that the trajectory-simulation model used here leads to predictions of mean concentration that are in very good agreement with atmospheric measurements of concentration obtained downwind of a continuous point source. There are no benchmark concentration data for an instantaneous point source release that can be used to compare with concentration estimates obtained using the fast kernel density estimator. In consequence, for the case of an instantaneous point source, the "true" mean concentration that will be used as a reference for comparison will be obtained using a very large number (5×10^6) particle trajectories with the "true" concentration obtained using the conventional box-counting method with small grid cells (cubes) of side length one meter.

Figures 6 and 7 exhibit streamwise (x-wise) cross-sections of the normalized mean concentration C/Q estimates at 6 different times after the release obtained using the fast kernel

density estimator. These streamwise (x-wise) cross-sections were obtained at $(y,z) = (1\text{ m}, 1\text{ m})$ and $(y,z) = (1\text{ m}, 3\text{ m})$, respectively. The estimates were obtained without applying the virtual (ghost) particle corrections near the ground surface as described in Eq. (25). The estimates used the parabolic and quadweight kernel functions with a constant bandwidth $h = 2$ m. The kernel density estimates were based on 50,000 independent particle trajectories. In spite of this, there is generally a good conformance of the quadweight kernel density estimate for the mean concentration with the "true" mean concentration based on 5×10^6 independent particle trajectories at both $z = 1$ and 3 m (cf. Figures 6 and 7, respectively). The kernel density estimator produces relatively "smooth" concentration estimates for even 50,000 particles. The parabolic kernel density estimator underestimated the "true" mean concentration at $z = 1$ m, although exhibited good conformance with the "true" mean concentration at $z = 3$ m. This is because the shape of parabolic kernel is broader [see Figure 1] than that of the quadweight kernel and at $z = 1$ m above the ground, the parabolic kernel function with $h = 2$ m has its domain of support "truncated" significantly by the ground surface at $z = 0$ m.

Figure 8 exhibits streamwise (x-wise) cross-sections of the normalized mean concentration C/Q estimates at $(y,z) = (1\text{ m}, 1\text{ m})$ and at 6 different times after the release, obtained using the fast kernel density estimator. However, these results were calculated using the kernel density estimator of Eq. (25) which accounts explicitly for the ground boundary effects by introducing for each particle at a vertical distance less than h (bandwidth) from the ground an associated virtual (ghost) particle below the ground surface placed at the mirror image location relative to the ground plane at $z = 0$. Note that the mean concentration estimates based on the parabolic kernel are greatly improved (cf. Figure 6) when the ground boundary effects are incorporated.

Figure 9 shows the simulation results for streamwise (x-wise) cross-sections of the normalized concentration C/Q estimates obtained at $(y,z) = (1\text{ m}, 3\text{ m})$ using a fast kernel density estimator (for both parabolic and quadweight kernel functions with $h = 2$ m). These estimates were obtained using only 20,000 independent particle trajectories. In spite of this, intercomparison of the kernel estimates with the "true" concentration shows good agreement, although not unexpectedly the statistical variability in the kernel estimates are larger than those exhibited in Figure 7 for 50,000 particles. Finally, Figure 10 displays the fast kernel density estimates of the normalized mean concentration C/Q obtained in streamwise (x-wise) cross-sections through the dispersing cloud at $(y,z) = (1\text{ m}, 3\text{ m})$. This kernel density estimate was obtained using only 10,000 independent particle trajectories for a parabolic kernel with a larger bandwidth $h = 6$ m. In comparison with Figure 9, note that increasing the bandwith h from 2 m to 6 m reduced the statistical error (variability) in the concentration estimate, but increased the bias. This bias is largest near the source release (cf. kernel-based concentration estimate at $t = 10$ s with the "true" concentration where it is seen that the peak mean concentration is underestimated, but the width of the streamwise concentration profile is overestimated). In this example, a smoothing length $h = 6$ m is too large and results in an over-smoothing of the mean concentration near the source (viz., at small travel times) where changes in the concentration are sharp. However, further from the source when the mean concentration gradients are not as large, the larger kernel bandwidth leads to good estimates

for the concentration with even 10,000 particle trajectories.

Conclusion

The kernel density estimator for the mean concentration field given information embedded in "marked" fluid particle trajectories provided an important improvement on previous methods for concentration estimation based on box-counting techniques that were sensitive to the size and location of imaginary sampling volumes. In spite of this, kernel estimators have rarely been used in Lagrangian particle modeling because the straightforward (direct) implementation of this method leads to a computational effort that is prohibitive as the computational work scales as $O(N^2)$ (where $N \approx N_r \approx N_p$). In this report, we present and implement an algorithm for the rapid evaluation of the kernel density estimate of the mean concentration field that requires an amount of computational effort proportional to N, and this method is insensitive to the precise distribution of the "marked" fluid particles in physical space (degree of disorder of the fluid particle positions which result from the stochastic trajectory-simulation model). In practice, speedups of two to three orders of magnitude may be expected in application of the fast kernel density estimator compared to the naïve approach, depending on the number of particles simulated and the number of receptor locations at which the mean concentration needs to be determined, rendering previously computationally prohibitive simulations feasible. In this report, both two- and three-dimensional versions of the fast kernel density estimator have been constructed and implemented. Results from a number of numerical experiments have been conducted to verify the algorithm.

There are several ways in which the code for the fast linked list based algorithm for the density kernel estimator of the mean concentration field can be either generalized and/or made more efficient. The linked list data structure used here is appropriate for the case where the bandwidth of the density kernel associated with each particle is fixed. In principle, the bandwidth h need not be constant and for many applications it may be useful to allow a variable smoothing length that is dynamically adapted so that the number of neighboring fluid particles within the domain of support of a kernel function centered on any particular particle remains constant. The generalization of the proposed fast kernel density estimator for variable kernel bandwidth h will probably require that the linked list data structure used here be replaced by a hierarchy tree data structure that can be adopted to suit the needs of a variable bandwidth. In the application of the fast kernel density estimator, the code can be made even more efficient by pre-computing the values of the kernel function at a large number of representative points and storing the results in a pre-computed kernel function value look-up table. Finally, it may well be worth exploring the parallelization of the code so that it can be executed on computers with highly parallel architectures (e.g., Beowulf clusters).

References

1. Thomson, D. J. (1987). Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of Fluid Mechanics*, 180, 529–556.

2. Du, S., Wilson, J. D. and Yee, E. (1994). Probability density functions for velocity in the convective boundary layer, and implied trajectory models. *Atmospheric Environment*, 28, 1211–1217.

3. Flesch, T. and Wilson, J. D. (1992). A two-dimensional trajectory-simulation model for non-Gaussian, inhomogeneous turbulence within plant canopies. *Boundary-Layer Meteorology*, 61, 349–374.

4. Luhar, A. K. and Britter, R. E. (1989). A random walk model for dispersion in inhomogeneous turbulence in a convective boundary layer. *Atmospheric Environment*, 23, 1911–1924.

5. Luhar, A. K. and Rao, K. S. (1994). Lagrangian stochastic dispersion model simulations of tracer data in nocturnal flows over complex terrain. *Atmospheric Environment*, 28, 3417–3431.

6. Yamada, T., Jim-Kao, C. H. and Bunker, S. (1989). Airflow and air quality simulations over the western mountainous region with a four-dimensional data assimilation technique. *Atmospheric Environment*, 23, 539–544.

7. Uliasz, M. (1994). Lagrangian particle dispersion modeling in mesoscale applications. In Zannetti, P., (Ed.), *Environmental Modeling*, Vol. 2, Southampton, UK: Computational Mechanics Publications.

8. de Haan, P. (1999). On the use of density kernels for concentration estimations with particle and puff dispersion models. *Atmospheric Environment*, 33, 2007–2021.

9. Batchelor, G. K. (1949). Diffusion in a field of homogeneous turbulence. Part I: eulerian analyis. *Australian Journal of Scientific Research*, 2, 437–450.

10. Du, S. (1997). Universality of the lagrangian velocity structure function constant (C_0) across different kinds of turbulence. *Boundary-Layer Meteorology*, 83, 207–219.

11. Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, London, UK: Chapman & Hall.

12. Epanechnikov, V. K. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Applications*, 14, 153–158.

13. Rodean, H. C. (1996). Stochastic Lagrangian Models of Turbulent Diffusion, Boston, MA: American Meteorological Society.

14. Barad, M. L. (1961). Project Prairie Grass, a field program in diffusion (Vol. II). (Technical Report 58-235). Air Force Cambridge Research Centre. Cambridge, MA.

15. Haugen, D. A. (1973). Workshop on Micrometeorology, Boston, MA: American Meteorological Society.
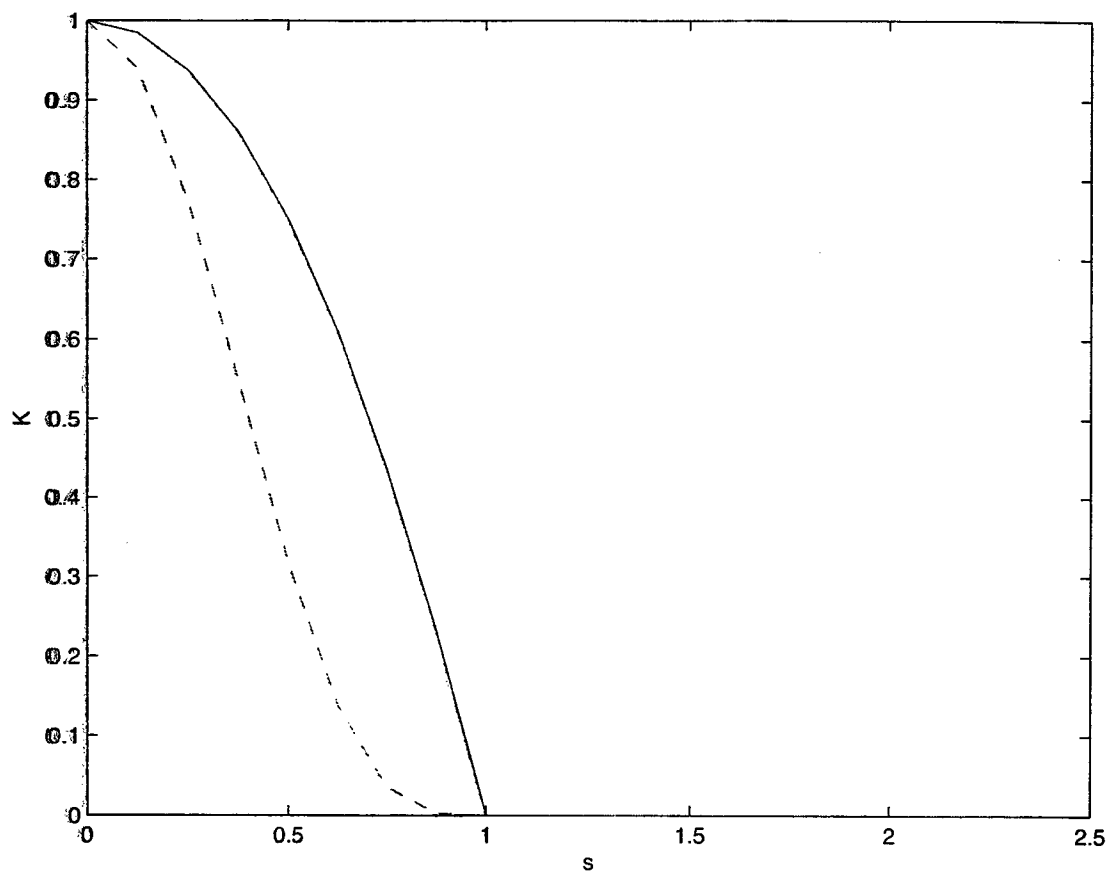
Figure 1: Profile of the non-normalized Epanechnikov (parabolic) (dashed line) and quadweight (solid line) kernels.
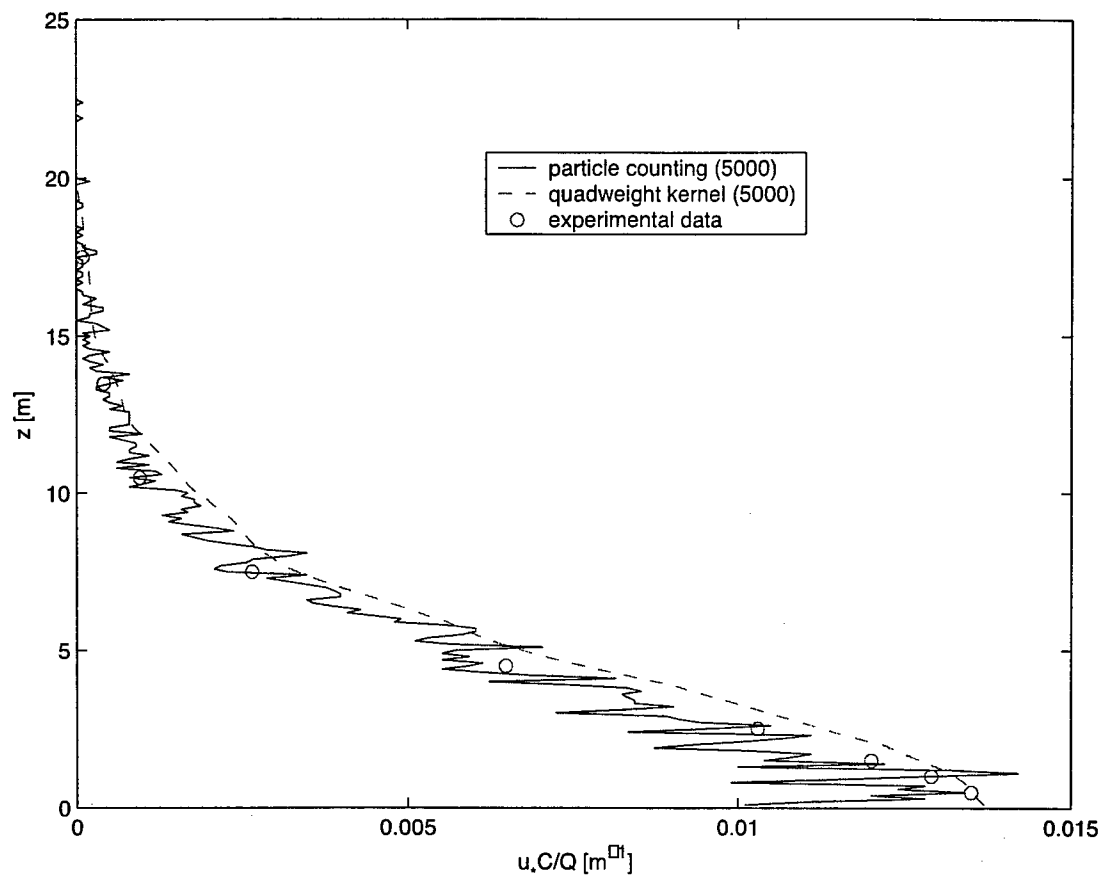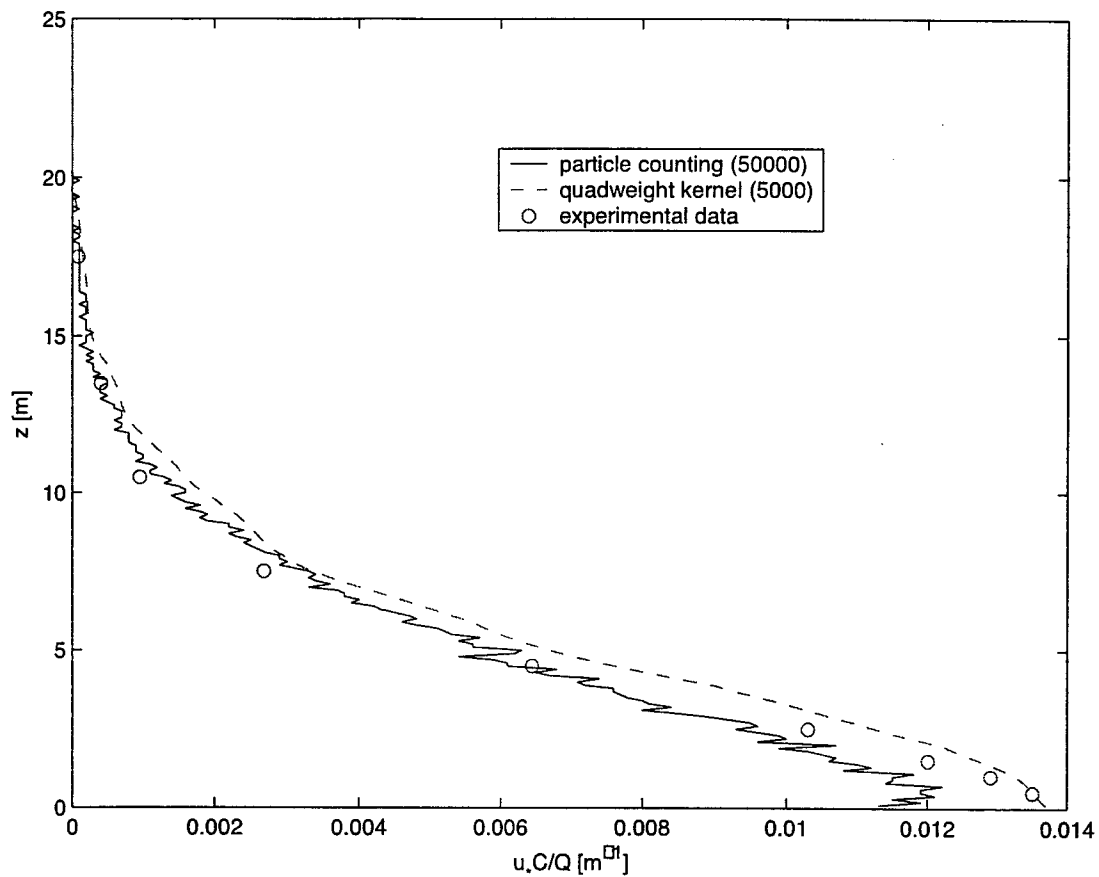
Figure 2: *Normalized mean concentration estimates based on the box-counting and fast kernel density estimator methods compared with observations obtained from the Project Praire Grass experiment. The concentration estimates were based on 5,000 independent particle trajectories calculated using a Lagrangian Stochastic model. The kernel function used for the mean concentration estimate is the quadweight kernel.*
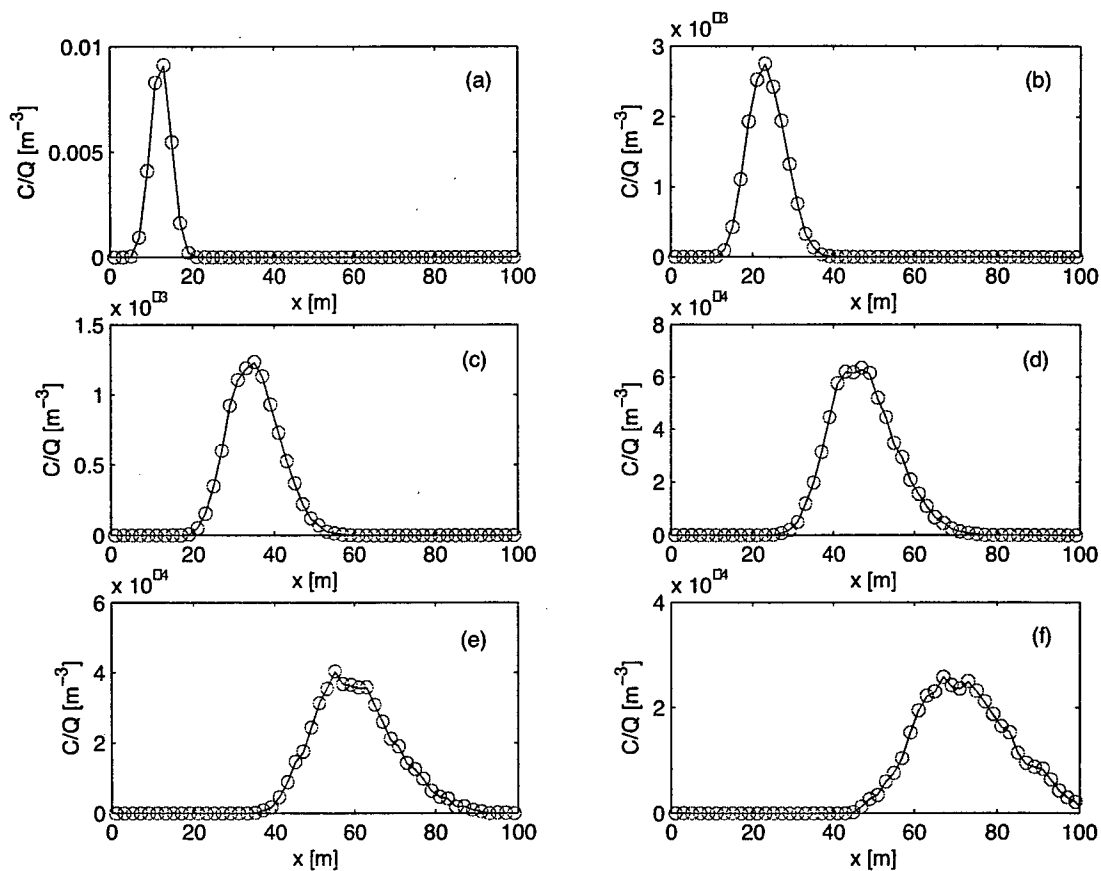
Figure 3: *Normalized mean concentration estimates based on the box-counting and fast kernel density estimator methods compared with observations obtained from the Project Praire Grass experiment. The concentration estimates obtained using the box-counting and fast kernel density estimator were based, respectively, on 50,000 and 5,000 independent particle trajectories calculated using a Lagrangian Stochastic model. The kernel function used for the mean concentration estimate is the quadweight kernel.*
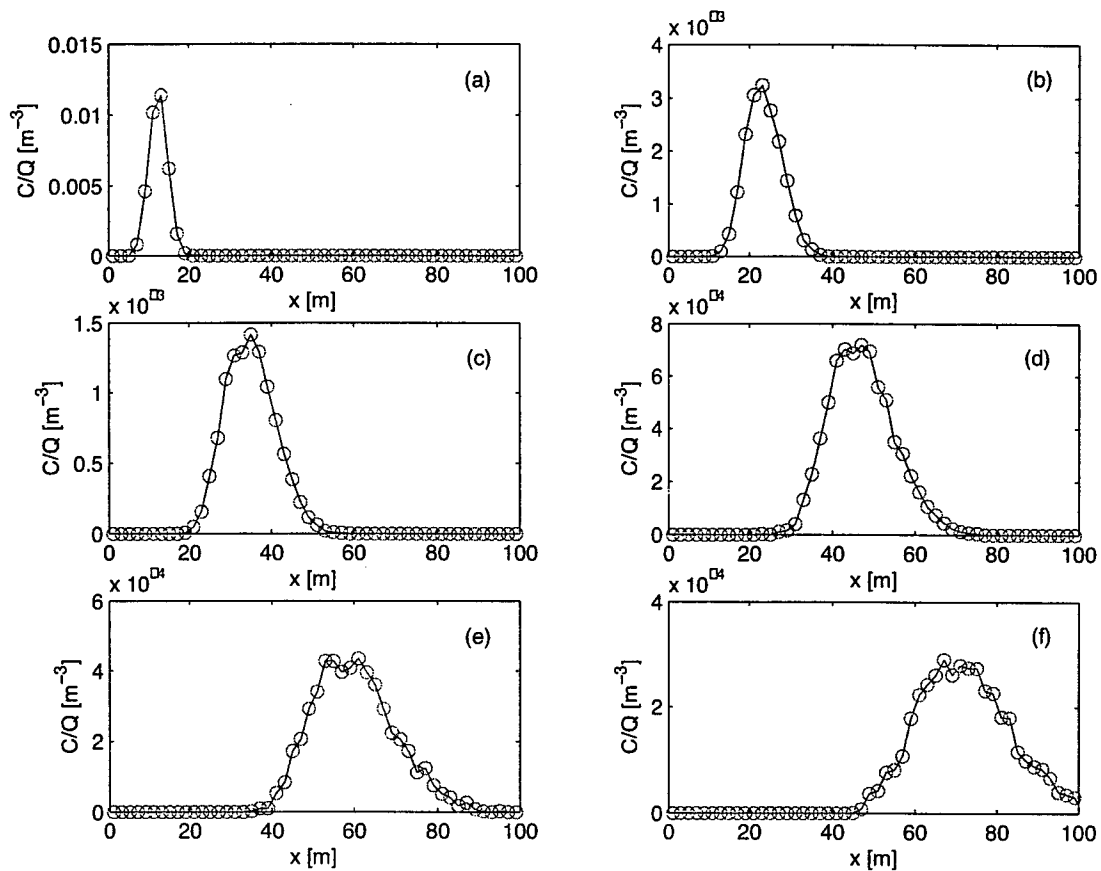
Figure 4: Simulation results for the normalized mean concentration along the mean cloud centerline at $y = 0$ and at height $z = 1$ m obtained using the direct and fast kernel density estimators for an instantaneous point source release at $x = y = 0$ and $z = 1$ m. The mean concentration is obtained at 6 different times after the tracer was released from the point source; namely, at (a) $t = 10$, (b) $t = 20$, (c) $t = 30$, (d) $t = 40$, (e) $t = 50$, and (f) $t = 60$ seconds. The kernel function used for this simulation is the parabolic (Epanechnikov) kernel with a bandwidth $h = 2$ m. The solid line and open circles represent the results obtained from the direct and fast kernel density estimators, respectively.
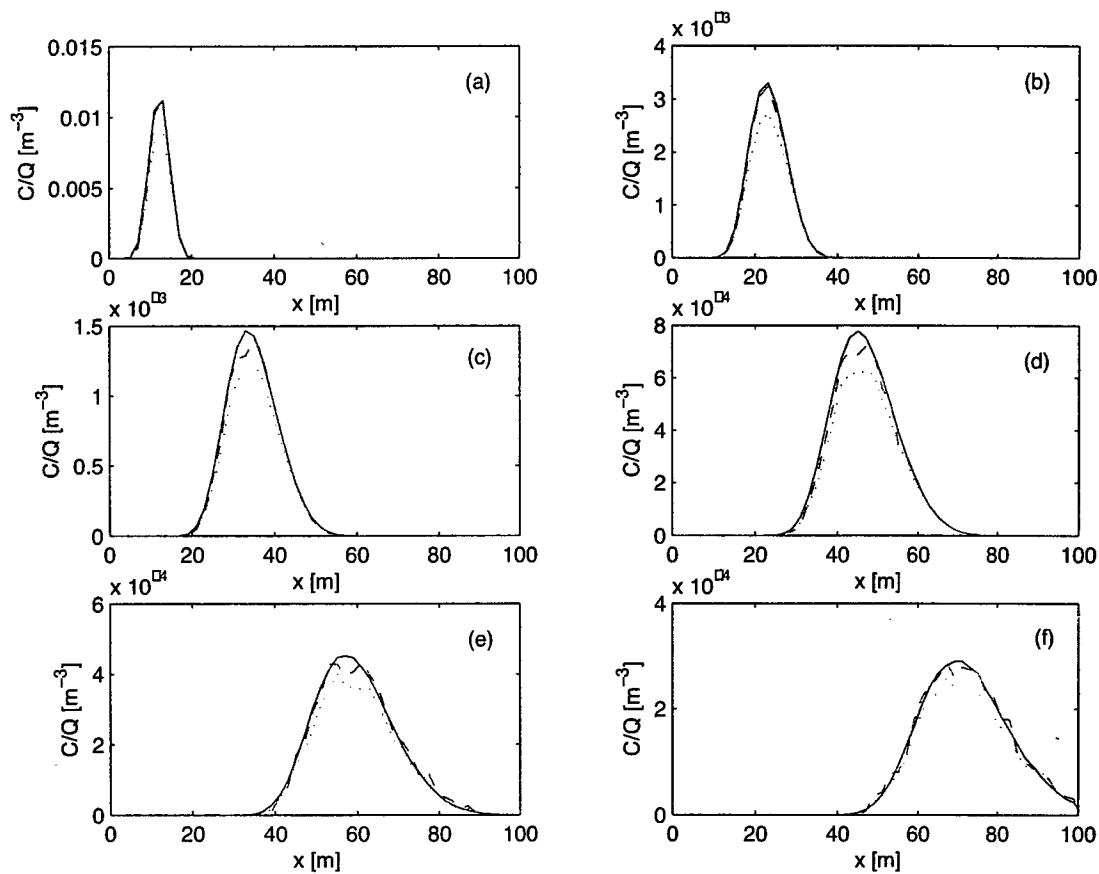
Figure 5: Simulation results for the normalized mean concentration C/Q along the mean cloud centerline at $y = 0$ and at height $z = 1$ m obtained using the direct and fast kernel density estimators for an instantaneous point source release at $x = y = 0$ and $z = 1$ m. The mean concentration is obtained at 6 different times after the tracer was released from the point source; namely, at (a) $t = 10$, (b) $t = 20$, (c) $t = 30$, (d) $t = 40$, (e) $t = 50$, and (f) $t = 60$ seconds. The kernel function used for this simulation is the quadweight kernel with a bandwidth $h = 2$ m. The solid line and open circles represent the results obtained from the direct and fast density kernel estimators, respectively.

Figure 6: Simulation results showing streamwise (x-wise) cross-sections of the normalized mean concentration C/Q obtained at $(y,z) = (1 \text{ m}, 1 \text{ m})$ for an instantaneous point source release at $x = y = 0$ and $z = 1$ m. The mean concentration is obtained at 6 different times after the tracer was released from the point source; namely, at (a) $t = 10$, (b) $t = 20$, (c) $t = 30$, (d) $t = 40$, (e) $t = 50$, and (f) $t = 60$ seconds. The normalized concentration estimates were obtained from 50,000 independent particle trajectories using the fast kernel density estimator with a parabolic (dotted line) and quadweight (dashed line) kernel functions with a constant bandwidth $h = 2$ m. The solid line corresponds to the "true" mean concentration obtained from 5×10^6 independent particle trajectories using a box-counting method. The effects of the ground surface on the kernel density estimate were not included in this simulation.
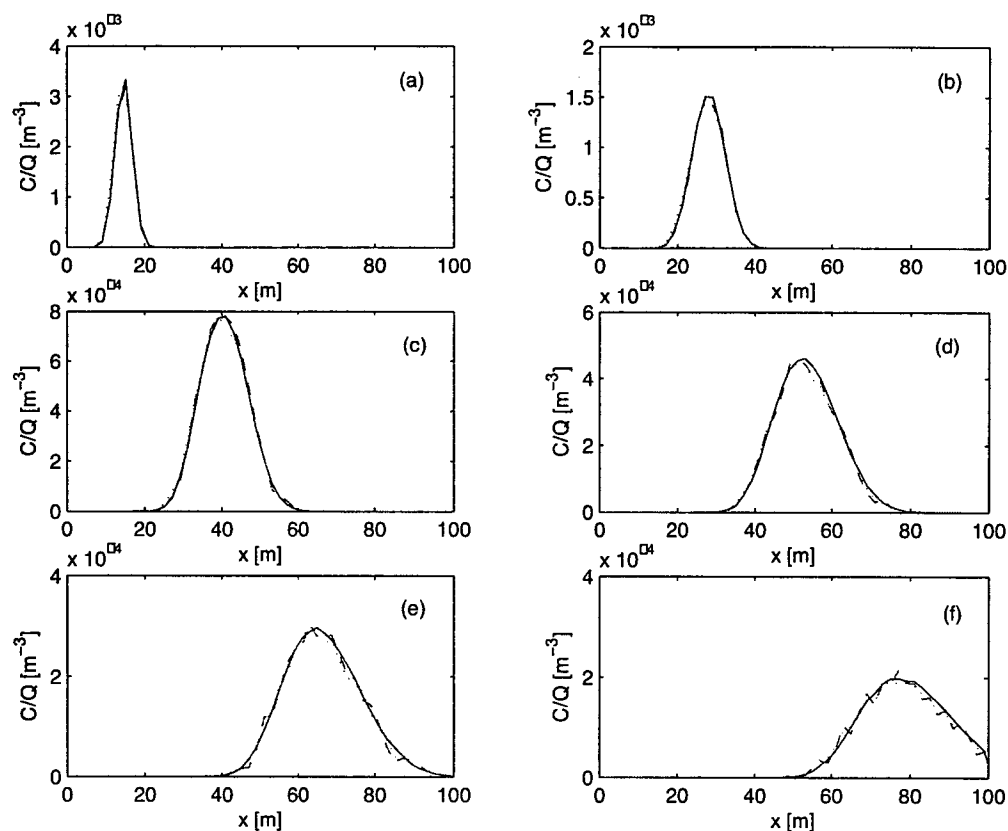
Figure 7: Simulation results showing streamwise (x-wise) cross-sections of the normalized mean concentration C/Q obtained at $(y,z) = (1\,\text{m}, 3\,\text{m})$ for an instantaneous point source release at $x = y = 0$ and $z = 1\,m$. The mean concentration is obtained at 6 different times after the tracer was released from the point source; namely, at (a) $t = 10$, (b) $t = 20$, (c) $t = 30$, (d) $t = 40$, (e) $t = 50$, and (f) $t = 60$ seconds. The normalized concentration estimates were obtained from 50,000 independent particle trajectories using the fast kernel estimator with parabolic (dotted line) and quadweight (dashed line) kernel functions with a constant bandwidth $h = 2\,m$. The solid line corresponds to the "true" mean concentration obtained from 5×10^6 independent particle trajectories using a box-counting method. The effects of the ground surface on the kernel density estimate were not included in this simulation.
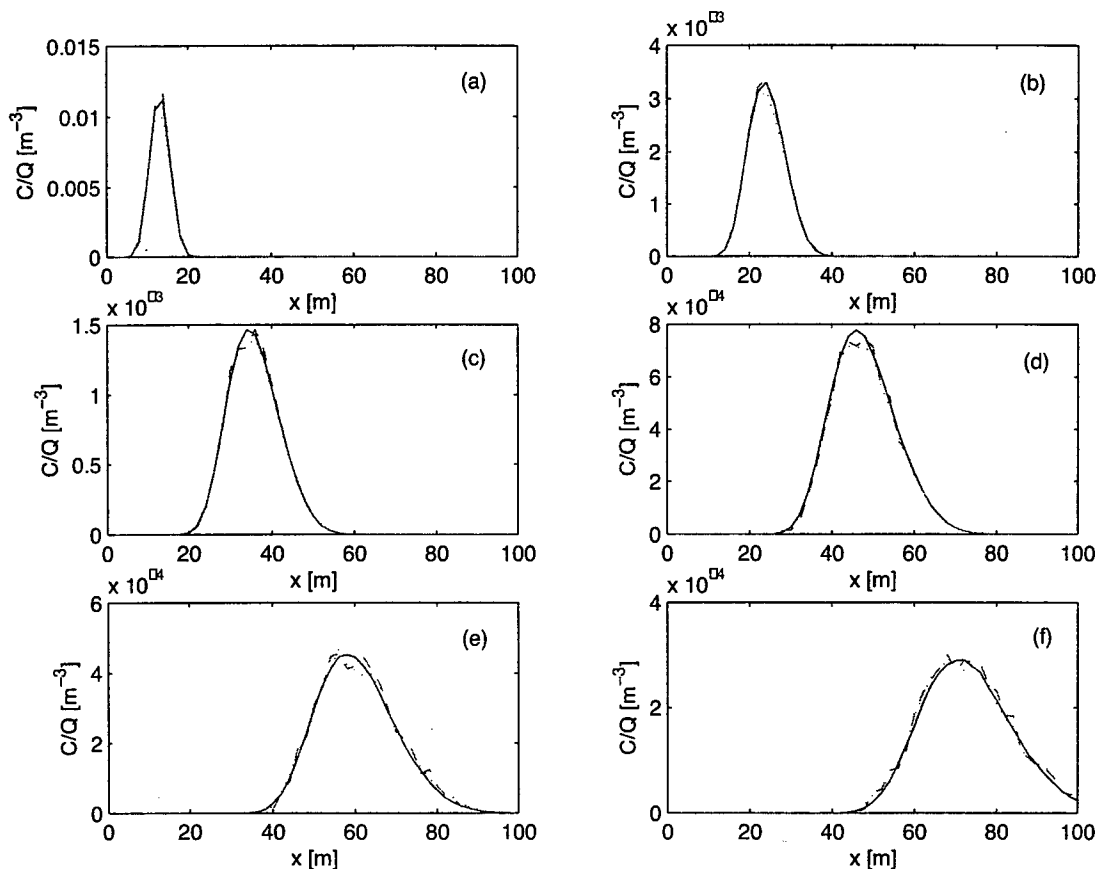
Figure 8: Simulation results showing streamwise (x-wise) cross-sections of the normalized mean concentration C/Q obtained at $(y, z) = (1 \text{ m}, 1 \text{ m})$ for an instantaneous point source release at $x = y = 0$ and $z = 1$ m. The mean concentration is obtained at 6 different times after the tracer was released from the point source; namely, at (a) $t = 10$, (b) $t = 20$, (c) $t = 30$, (d) $t = 40$, (e) $t = 50$, and (f) $t = 60$ seconds. The normalized concentration estimates were obtained from 50,000 independent particle trajectories using the fast kernel density estimator with parabolic (dotted line) and quadweight (dashed line) kernel functions with a constant bandwidth $h = 2$ m. The solid line corresponds to the "true" mean concentration obtained from 5×10^6 independent particle trajectories using a box-counting method. The effects of the ground surface on the kernel density estimate were included in this simulation.
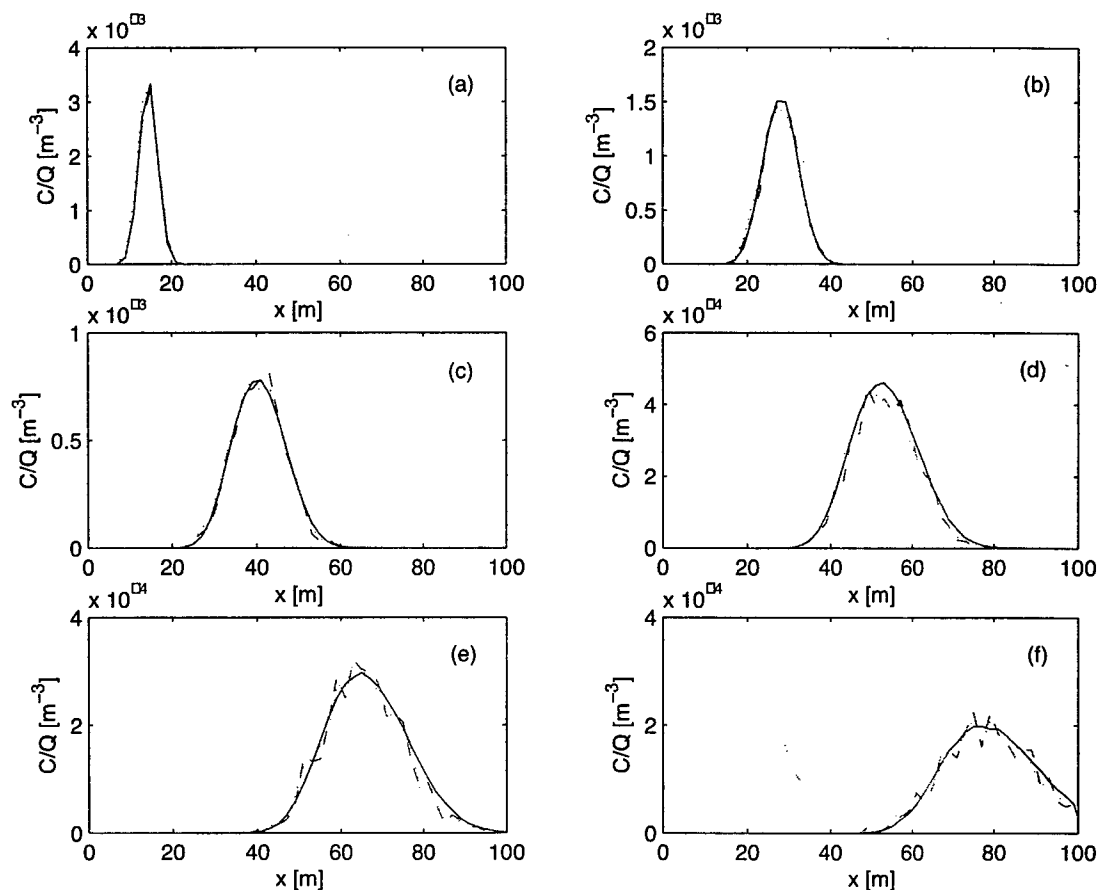
Figure 9: Simulation results showing streamwise (x-wise) cross-sections of the normalized mean concentration C/Q obtained at $(y,z) = (1 \text{ m}, 3 \text{ m})$ for an instantaneous point source release at $x = y = 0$ and $z = 1$ m. The mean concentration is obtained at 6 different times after the tracer was released from the point source; namely, at (a) $t = 10$, (b) $t = 20$, (c) $t = 30$, (d) $t = 40$, (e) $t = 50$, and (f) $t = 60$ seconds. The normalized concentration estimates were obtained from 20,000 independent particle trajectories using the fast kernel density estimator with parabolic (dotted line) and quadweight (dashed line) kernel functions with a constant bandwidth $h = 2$ m. The solid line corresponds to the "true" mean concentration obtained from 5×10^6 independent particle trajectories using a box-counting method.
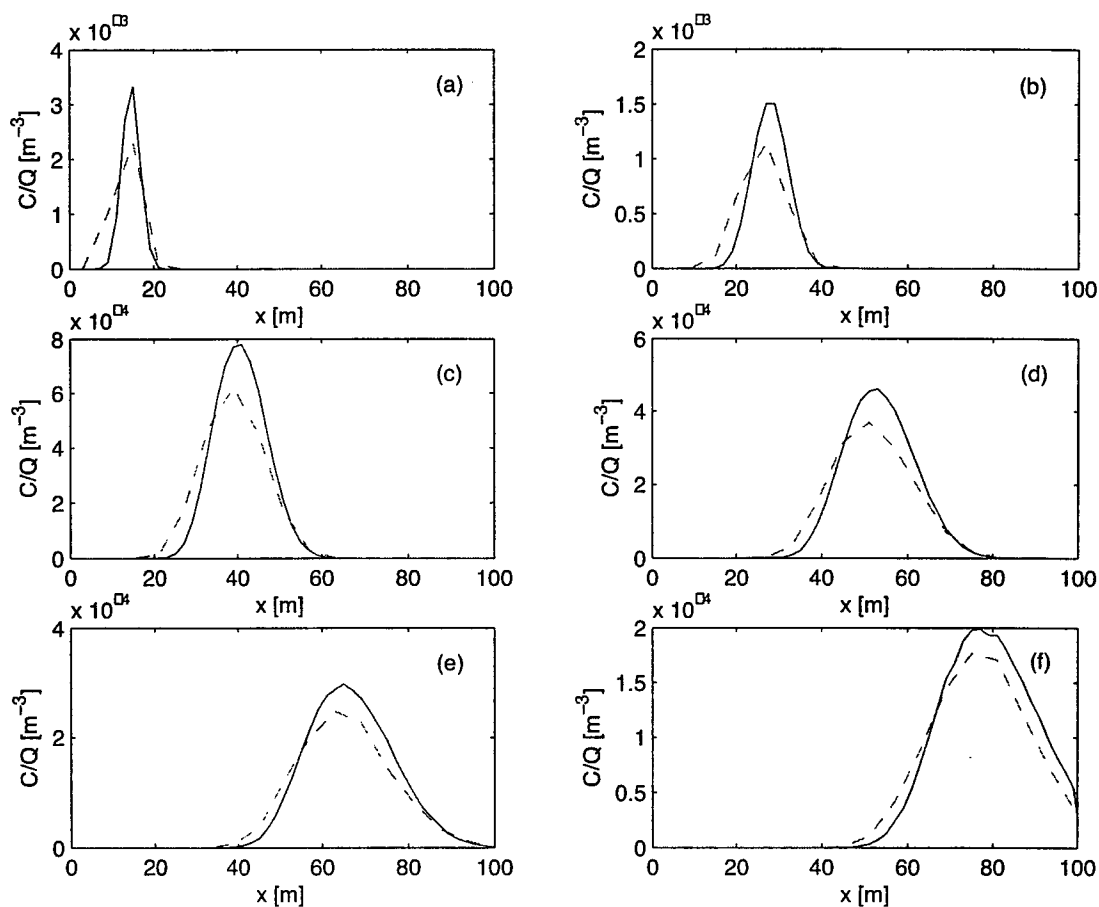
Figure 10: Simulation results showing streamwise (x-wise) cross-sections of the normalized mean concentration C/Q obtained at $(y,z) = (1 \text{ m}, 3 \text{ m})$ for an instantaneous point source release at $x = y = 0$ and $z = 1$ m. The mean concentration is obtained at 6 different times after the tracer was released from the point source; namely, at (a) $t = 10$, (b) $t = 20$, (c) $t = 30$, (d) $t = 40$, (e) $t = 50$, and (f) $t = 60$ seconds. The normalized concentration estimates were obtained from 10,000 independent particle trajectories using the fast kernel density estimator with a quadweight (dashed line) kernel function with a constant bandwidth $h = 6$ m. The solid line corresponds to the "true" mean concentration obtained from 5×10^6 independent particle trajectories using a box-counting method.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R & D Canada – Suffield PO Box 4000, Medicine Hat, AB, Canada T1A 8K6	2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable). UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title). The Rapid Evaluation of Mean Concentration Fields in Lagrangian Stochastic Modelling Using a Density Kernel Estimator

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.) E., Shao, Y. and Yee,

5. DATE OF PUBLICATION (month and year of publication of document) October 2004	6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc). 41	6b. NO. OF REFS (total cited in document) 15

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered). Technical Report

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include address). Defence R & D Canada – Suffield PO Box 4000, Medicine Hat, AB, Canada T1A 8K6

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Specify whether project or grant). PCN 16QD11	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written).

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique.) DRDC Suffield TR 2004-186	10b. OTHER DOCUMENT NOs. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification) (X) Unlimited distribution ( ) Defence departments and defence contractors; further distribution only as approved ( ) Defence departments and Canadian defence contractors; further distribution only as approved ( ) Government departments and agencies; further distribution only as approved ( ) Defence departments; further distribution only as approved ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution beyond the audience specified in (11) is possible, a wider announcement audience may be selected).

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

Lagrangian Stochastic (LS) particle models have proven to be a useful computational tool for the description and prediction of dispersion of pollutant releases in complex meteorological situations (e.g., space- and time-varying situations pertaining to complex flow and turbulence). However, simulating the emitted pollutant by following the trajectories of many "marked" fluid elements released from the source distribution brings up the difficulty of the correct estimation of the mean concentration of the dispersing pollutant from the particle trajectory information. Recently, the density kernel estimation method has been proposed and applied successfully to estimate mean concentrations from Lagrangian Stochastic particle models. However, the computational effort needed by this method increases as N^2 (assuming the number of receptor locations N_r at which the concentration is required is comparable to the number of fluid particles N_p used in the trajectory simulation, so $N_r \approx N_p \sim N$) and, in consequence, the method has not been widely used because of the significant computer resources required. Here, we describe a novel algorithm for calculating the kernel estimate of the mean concentration field whose computational complexity scales only as N. The technique uses a tesselation (subdivision) of space in cubic cells of side length h (where h is the bandwidth of the kernel function), and then associates a linked-list data structure with each cell that is used as a bookkeeping device to keep track of the "marked" fluid particles in that cell. The fast approach developed here has been verified by comparing results with the direct implementation of the kernel estimator and with the conventional box-counting estimator for the mean concentration field.