Atmospheric Environment 43 (2009) 3691–3698

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

The kernel method to compute the intensity of segregation for reactive pollutants: Mathematical formulation

Gianni Pagnini*

ENEA, ACS PROT-INN, via Martiri di Monte Sole 4, 40129 Bologna, Italy

ARTICLE INFO

Article history:
Received 31 July 2008
Received in revised form
21 April 2009
Accepted 21 April 2009

Keywords:
Chemically reactive plume
Mixing of chemical species
Intensity of segregation
Effective reaction rate
Kernel method

ABSTRACT

It is well known that turbulent dispersion influences chemical reactions and that computation of reactant concentrations or mean chemical reaction rates can suffer of serious error when small-scale atmospheric processes' effects on chemical transformation are neglected. A quantity that gives a measure of the influence of turbulent dispersion on second-order chemical reaction rates is the *intensity of segregation*. A nonparametric estimator based on the kernel method aimed at measuring the intensity of segregation is proposed. Numerical benchmark tests, in the case of a Gaussian plume, are performed to study the suitability of this technique. The estimator works well, especially for small and moderate separation from the plume centreline and generally in the smooth parts of the estimated function. The effective reaction rate is computed and the percentage error emerges to be less than 5% in the best estimation intervals, and less than 40% in the worst. A method to reduce percentage error is introduced and improved performances are observed. The estimator proposed turns out to be particularly suitable for Lagrangian air quality modelling because it permits conservation of the grid independence.

1. Introduction

The present study is motivated by the importance to understand the effects of turbulent mixing on atmospheric chemical reactions. This subject, besides theoretical aspects, is important for environmental problems, e.g. air quality modelling.

The atmosphere is turbulent and turbulent dispersion drives the mixing of reactive scalars. It is well known that the "goodness of mixing" is a fundamental factor in chemical reactions and that the optimal condition for chemical reactions to act is the uniform distribution of reactants. This "goodness of mixing" is measured by a statistical quantity called *intensity of segregation* (Danckwerts, 1952). Inhomogeneous mixing due to turbulence generates the segregation of species, and its role is not negligible when chemical time-scales are comparable or less than the turbulent time-scale. The ratio of turbulent and chemical time-scales is called the Damköhler number. In this case turbulent motion influences the mixing of reactants before they react (Vinuesa and de Arellano, 2005). Under typical atmospheric conditions, the reactions of hydroxyl radicals and the reactions of peroxy radicals with NO are influenced by turbulence. Several second-order chemical reactions affected by atmospheric turbulence are given by Stockwell (1995). The segregation of species can cause a decrease of reaction rate if reactants are non-premixed or an increase if they are premixed (see e.g. duP. Donaldson and Hilst, 1972; Galmarini et al., 1995; Molemaker and de Arellano, 1998; de Arellano, 2003; de Arellano et al., 2004; Vinuesa and de Arellano, 2005). Generally, in air quality problems, reactants are non-premixed because they can come from different sources or because one is emitted from a source and the other is already present in the environment. Here only the non-premixed case is considered.

Numerical Eulerian atmospheric models divide the spatial domain in boxes and the governing equations are solved in the points of this grid. In the case of a reactive plume, the larger spatial scale of the intensity of segregation is the mean plume size. Generally, after the emission, instantaneous and homogeneous mixing of reactants is assumed. This artificial dilution leads to an initial underestimation of NO_x and other primary emitted species and overestimation of secondary products such as O_3, HO and PAN. For these reactions the model accuracy is strongly limited and a method to include chemical segregation is needed (Molemaker and de Arellano, 1998; Vinuesa and de Arellano, 2005). Serious errors can be made in estimation of reactant concentrations (de Arellano, 2003), or mean chemical reaction rates (Komori et al., 1991), when small-scale atmospheric process effects on chemical transformations are neglected. From above analysis emerges that the subgrid problem arises to model turbulence–chemistry interrelation expressed by the intensity of segregation. Further investigations on this topic are necessary.

* Present address: CRS4, Polaris Bldg. 1, 09010 Pula (CA), Italy. Tel.: +39 0709250391; fax: +39 0709250216.
E-mail address: pagnini@crs4.it

1352-2310/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2009.04.049

The analytical solutions of the advection–reaction–diffusion equation can be adopted, but to find solutions for general situations is a very difficult task, both for first (Cokca, 2003) and second-order reactions (Rubio et al., 2008).

It is well known that besides Eulerian models there are Lagrangian models. The Lagrangian approach permits to study dispersion in very general circumstances, both in respect of atmospheric stability and source types. However, in many cases, in Lagrangian framework the concentration fields are estimated by counting the number of particles in a rectangular volume (box counting) and the choice of the width and the position of these boxes influences the computation (de Haan, 1999). Recently, in order to have totally grid-free Lagrangian models, successful applications of kernel method for concentration estimations appeared in literature, e.g. (Lorimer, 1986; Lorimer and Ross, 1986; de Haan, 1999; Davakis et al., 2003; Vitali et al., 2006; Monforti et al., 2006).

The kernel method is a nonparametrical technique where, differently from classical parametric approach, no assumptions are made on the statistics of the underlying data. The principal aim of the present analysis is to study the suitability of the kernel method to estimate the intensity of segregation in chemical reactions in inhomogeneous mixing, in particular its profile inside a reactive plume/puff.

Concerning Lagrangian modelling, Gaussian or analytical models, which give formulae for the concentration field only in very simplified situations, have been substituted by complex stochastic models, which give the opportunity to study dispersion in several different conditions even if the concentration field is not analytically computable. The development of the kernel method, in order to compute the concentration fields produced by complex Lagrangian models, permits the analysis of pollutant dispersion in real situations (e.g. Davakis et al., 2003). In analogy with this, in the present paper I do not propose an approximated analytical model or some parameterizations of the intensity of segregation that hold only in some particular cases, but a statistical method to compute the intensity of segregation generated by complex Lagrangian models in real situations. In the same spirit of concentration field computation, the new estimator is straightforwardly applicable in Lagrangian air quality modelling with any type of source. It overcomes the grid-size problem because it is a grid-free approach. This opportunity to measure the intensity of segregation of a complex Lagrangian model with grid independency is the key and novel aspect of the present research.

The rest of the paper is organized as follows. In Section 2 the background and some previous approaches are briefly reviewed. In Section 3 the mathematical formulation of the nonparametric estimation proposed is derived and a couple of remarks on it are highlighted. In Section 4 numerical tests to study the efficiency of the estimations are performed in order to reproduce the intensity of segregation function obtained by Galmarini et al. (1995) with a Gaussian reactive plume. In Section 5 the results are discussed and the conclusions are given.

2. Background and some previous approaches

Let α and β be two reactive species so that $\alpha+\beta \xrightarrow{k}$ products, where k is the reaction rate. Let $c = C + c'$ where c is the concentration, C the average concentration and c' the concentration fluctuation of a generic reactive scalar. The instantaneous concentration of α is given by

$$\frac{\partial c_\alpha}{\partial t}+\frac{\partial}{\partial x_i}(u_i c_\alpha)-D_\alpha\frac{\partial^2 c_\alpha}{\partial x_i \partial x_i} = -k c_\alpha c_\beta, \tag{1}$$

where t, $\mathbf{x}\in R^3$, $\mathbf{u}\in R^3$ and D_α are time, space, velocity field and molecular diffusivity of α, respectively. In the case of an incompressible flow, from averaging it follows that

$$\begin{aligned}\frac{\partial C_\alpha}{\partial t}+U_i\frac{\partial C_\alpha}{\partial x_i}+\frac{\partial}{\partial x_i}\langle u_i' c_\alpha'\rangle - D_\alpha\frac{\partial^2 C_\alpha}{\partial x_i\partial x_i} &= -k\left[C_\alpha C_\beta+\left\langle c_\alpha' c_\beta'\right\rangle\right]\\ &= -k[1+I_S]C_\alpha C_\beta\\ &= -k_{eff}C_\alpha C_\beta,\end{aligned} \tag{2}$$

where I_S and k_{eff} are the *intensity of segregation* and the *effective reaction rate*, respectively, and they are defined as

$$I_S = \frac{\left\langle c_\alpha' c_\beta'\right\rangle}{C_\alpha C_\beta} = \frac{R'_{\alpha\beta}}{C_\alpha C_\beta} = \frac{R_{\alpha\beta}}{C_\alpha C_\beta}-1, \quad k_{eff} = k[1+I_S], \tag{3}$$

$R_{\alpha\beta}=\langle c_\alpha(\mathbf{x})c_\beta(\mathbf{x})\rangle$ and $R'_{\alpha\beta}(\mathbf{x})=\langle c'_\alpha(\mathbf{x})c'_\beta(\mathbf{x})\rangle$ are the correlations of concentration and concentration fluctuations, respectively, and they are related by $R_{\alpha\beta}=R'_{\alpha\beta}+C_\alpha C_\beta$. The same is for the reactant β. The lower bound of I_S is -1, when $\partial C/\partial t=0$, and I_S becomes positive only when the species are premixed before being introduced in the fluid flow (de Arellano, 2003). Because of here only non-premixed reactants are considered, hereinafter $-1\le I_S\le 0$. A non-null segregation means a non-uniform distribution of chemical species and then a concentration gradient. In inhomogeneous mixtures, the reaction rate is not a constant but it depends on the level of the "goodness of mixing" and the intensity of segregation is a statistical measure of this "goodness" (Danckwerts, 1952; duP. Donaldson and Hilst, 1972).

The first work on the statistical theory of turbulent chemical reactions is the one by Corrsin (1958). For further theoretical developments see Hill (1976), Pope (1985), Dopazo et al. (1997) and for models of second-order chemical reactions in turbulent flows Lamb and Shu (1978), Shu et al. (1978), Crone et al. (1999), van Dop (2001).

From the statistical theory of marked particles, if they are assumed to be one-chemical component, the average concentration and mean-square concentration are (Komori et al., 1991)

$$\langle c(\mathbf{x},t)\rangle = C(\mathbf{x},t) = \int p_1(\mathbf{x};t|\mathbf{x}_0,0)S(\mathbf{x}_0)d\mathbf{x}_0, \tag{4}$$

$$\begin{aligned}\left\langle c^2(\mathbf{x},t)\right\rangle &= \left\langle c^{(1)}(\mathbf{x},t)c^{(2)}(\mathbf{x},t)\right\rangle\\ &= \int p_2(\mathbf{x},\mathbf{x};t|\mathbf{x}_0^{(1)},\mathbf{x}_0^{(2)},0)S^{(1)}(\mathbf{x}_0^{(1)})S^{(2)}(\mathbf{x}_0^{(2)})d\mathbf{x}_0^{(1)}d\mathbf{x}_0^{(2)},\end{aligned} \tag{5}$$

where p_1 and p_2 are the Lagrangian density function for one particle and two particles, respectively, $S^{(i)}(\mathbf{x}_0^{(i)})$ is the source distribution of particle i at time $t=0$ and position $\mathbf{x}^{(i)}=\mathbf{x}_0^{(i)}$. If the reactants α and β are initially non-premixed

$$\begin{aligned}R_{\alpha\beta}(\mathbf{x},t) &= \left\langle c_\alpha(\mathbf{x},t)c_\beta(\mathbf{x},t)\right\rangle\\ &= \frac{1}{2}\left[\left\langle c_\alpha^{(1)}(\mathbf{x},t)c_\beta^{(2)}(\mathbf{x},t)\right\rangle+\left\langle c_\alpha^{(2)}(\mathbf{x},t)c_\beta^{(1)}(\mathbf{x},t)\right\rangle\right].\end{aligned} \tag{6}$$

Unfortunately, there is no analytically or computationally efficient method for obtaining exact solutions of p_1 and p_2.

There are substantially two research lines in literature: one focused on the segregation profile in the whole Convective Boundary Layer (CBL) and the other on its profile inside a reactive plume. For the first research line there is a great number of papers in literature, the CBL capacity to generate segregation is studied mainly by Large Eddy Simulation (LES) of bottom-up and top-down diffusing reactants. The first calculation of I_S for the CBL using LES is given by Schumann (1989). In his simulation subgrid-scale contributions of reactions are neglected. Similar simulations are performed for example in Sykes et al. (1994) and Molemaker and de Arellano (1998). Besides, Krol et al. (2000) extend the study to a more complex chemistry scheme that leads to substantially

different findings. Generally, LES is a valuable calculation of the atmospheric boundary layer characteristics and its results can be used to propose parameterizations (Petersen and Holtslag, 1999; Vinuesa and de Arellano, 2003, 2005). Analogously, in Sykes et al. (1994) the full closure equations for the second-moment scalar species correlation are derived and some model of the triple fluctuation correlation in terms of the first and second moments are prescribed and compare with LES data. Third-order models are derived also in Hilst (1998).

The second research line is much more poor of literature. This poorness follows, clearly, from grid resolution problems. Sykes et al. (1992) performed the first LES of a turbulent reactive plume. However, they assume a well-mixing inside the plume and found an approximately constant value of I_S over a large part of the plume cross-section. However, the prominent role played by the subgrid-scale mixing for a reactive plume into a neutral atmospheric boundary layer is shown in LES results by Meeder and Nieuwstadt (2000), where the intensity of segregation at the plume centreline is plotted as a function of downwind distance. In this respect the paper by Galmarini et al. (1995) is interesting. They used a Gaussian dispersion process that corresponds to a neutral atmospheric boundary layer and, using the *K*-theory, derived the concentration fluctuation correlation as proportional to the product of the gradient of mean concentration of both reacting species. The proportionality coefficient is dependent on the turbulent diffusivity *K*, the Damköhler number and some experimentally determined constants. They found that I_S inside the plume is dependent on the distance from the plume centreline. A similar closure is derived in Aguirre et al. (2005) where the dispersion process is driven by a LES coupled with a subgrid Lagrangian stochastic model and the chemical mechanism is evolved in time dividing the domain in boxes, selecting randomly particle pairs and then evolving particle concentration. In this last paper, the proportionality coefficient is chosen in order to obtain the best fit. In Karamchandani et al. (2000) the results of a Gaussian based dispersion model with a chemical mechanism and a model of the intensity of segregation are shown. However, no study of the variation of I_S inside the plume cross-section is shown. To conclude, other approaches are the so-called Conditional Moment Closure (CMC) (Klimenko and Bilger, 1999) and the Stochastic Field method (Valiño, 1998). This last is used in (Garmory et al., 2006) to study the micromixing effects in a reacting plume. The equation includes instantaneous chemical reaction rates so that no concentration fluctuation correlation is prescribed.

Finally, in all the briefly described approaches the Eulerian equations of the velocity field are closed using the *K*-theory and some criticism on using this theory in diffusion problems involving nonlinear chemical reactions is discussed by Lamb (1973), in particular in the vicinity of strong, localized sources such as smoke-stacks, highways etc.. Promising indications emerged from all approaches, but in certain cases, grid resolution and subgrid assumptions have some effects on the model prediction, as shown in (Chock et al., 2002) for a reactive plume under convective conditions.

In order to overcame this problem, in what follows a grid independent estimator based on kernel method aimed at measuring the intensity of segregation is proposed. Moreover, this independence makes the method particularly useful for analysing data obtained by Lagrangian modelling, which is the most suitable approach for studying the small-scale processes considered here.

3. Nonparametric estimators

The average concentrations $C_\alpha(\mathbf{x})$ and $C_\beta(\mathbf{x})$ and the concentration correlation $R_{\alpha\beta}(\mathbf{x})$ can be nonparametrically estimated using the kernel method (Hall and Patil, 1994; Hall et al., 1994; Pope, 2000). Let the symbol ^ indicate the nonparametric estimator, then for an ensemble of *N* particles, each located in $\mathbf{x}_i$ with $i = 1, \ldots, N$,

$$\widehat{C}(\mathbf{x}) = \frac{\sum_{i=1}^{N} c(\mathbf{x}_i) K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)}, \tag{7}$$

$$\widehat{R}_{\alpha\beta}(\mathbf{x}) = \frac{\sum_{i=1}^{N} c_\alpha(\mathbf{x}_i) c_\beta(\mathbf{x}_i) K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)}, \tag{8}$$

where, in (7), $c(\mathbf{x}_i)$ stands for $c_\alpha(\mathbf{x}_i)$ (or $c_\beta(\mathbf{x}_i)$) and it is the value assumed by the concentration field of the chemical compound α (or β) in the point $\mathbf{x}_i$, and it is used to compute $\widehat{C}_\alpha(\mathbf{x})$ (or $\widehat{C}_\beta(\mathbf{x})$) that is the estimation of the average concentration $C_\alpha(\mathbf{x})$ (or $C_\beta(\mathbf{x})$). In the present study the values of $c(\mathbf{x}_i)$ are assumed given and only their statistical analysis is considered. However, they can be obtained, for example, as the outputs of the concentration field in the points $\mathbf{x}_i$ computed by an air quality model. In both (7) and (8), $h(N) > 0$ is the bandwidth parameter, $h(N) \to 0$ as $N \to \infty$, and *K* is a *d*-variate kernel function so that $K(\xi) \geq 0$, $\int K(\xi) d\xi = 1$ and $\int K^2(\xi) d\xi < \infty$.

It is easy to recognize in (7) and (8) the usual weighted mean formulae, where the kernel function *K* is the weight of each *i*-th datum and it is based on the separation $\mathbf{x} - \mathbf{x}_i$. More, in the case of a uniform concentration distribution the weight function becomes a constant, each point has the same importance $K = const.$, and formula $\sum_{i=1}^{N} \cdot / N$ is recovered.

Finally, from (7) and (8), the nonparametric estimators of k_{eff} and I_S are

$$\widehat{k}_{eff} = k\left[1 + \widehat{I}_S\right], \quad \widehat{I}_S = \frac{\widehat{R}'_{\alpha\beta}}{\widehat{C}_\alpha \widehat{C}_\beta} = \frac{\widehat{R}_{\alpha\beta}}{\widehat{C}_\alpha \widehat{C}_\beta} - 1. \tag{9}$$

3.1. Remarks on $\widehat{R}_{\alpha\beta}(\mathbf{x})$

The estimator $\widehat{R}_{\alpha\beta}(\mathbf{x})$ defined in (8) requires some remarks.

For very large values of $x = |\mathbf{x}|$ the estimator $\widehat{R}_{\alpha\beta}$ can become undefined. In fact, the kernel function *K* goes to zero for any $h > 0$ and sufficiently large *x*, then the ratio in (8) becomes 0/0. Moreover, even if well-defined, the estimator $\widehat{R}_{\alpha\beta}$ can be quite accurate for small and moderate values of *x* but highly inaccurate for large *x*. This last problem can be solved for a monotonic decreasing correlation function stating its value equals to zero for $x > x_c$ ($x_c > 0$) (Hall and Patil, 1994). However, in the present application this problem cannot be solved in this simple way. In fact, the correlation function is supposed to decrease from 0 to a negative value ≥ -1 and then to grow up to 0. To establish the limit-value x_c is more difficult.

Moreover, when the function to estimate is not monotonic another problem arises around its maximum/minimum points. In fact, in the one-dimensional case, deriving in space formula (8) gives

$$\widehat{R}_{\alpha\beta}(x) = \frac{\sum_{i=1}^{N} c_\alpha(x_i) c_\beta(x_i) \frac{dK(x-x_i)}{dx}}{\sum_{i=1}^{N} \frac{dK(x-x_i)}{dx}} - \frac{d\widehat{R}_{\alpha\beta}}{dx} \frac{\sum_{i=1}^{N} K(x-x_i)}{\sum_{i=1}^{N} \frac{dK(x-x_i)}{dx}}. \tag{10}$$

Let $x = x_M$ be the maximum/minimum point, then from (10) it follows that

$$\widehat{R}_{\alpha\beta}(x_M) = \frac{\sum_{i=1}^{N} c_\alpha(x_i) c_\beta(x_i) \frac{dK(x-x_i)}{dx}\Big|_{x=x_M}}{\sum_{i=1}^{N} \frac{dK(x-x_i)}{dx}\Big|_{x=x_M}}. \tag{11}$$

Now, by definition the kernel functions $K(x)$ are symmetrical probability density functions with zero mean. From this it follows

that the maximum, and then the mean, of functions $K(x - x_i)$ is located at $x = x_i$, as one expects from a weighted mean interpretation in order to give more importance to data close to the point under consideration, i.e. x. But it follows also that $dK(x - x_i)/dx = 0$ in $x = x_i$, and this brings some consequence. In fact, in formula (11) the non-zero addenda, in both sums, are only those with $x_i \neq x_M$, that are the less significant data because they are not located in the point under consideration, i.e. $x = x_M$. This remains true for a set of data located in $x_i \simeq x_M$. Then, a reduction of performance of the estimation around the maximum/minimum point $x = x_M$ is expected, because the most significant data are lost. So, even if in general one expects that the enlarging of N improves the estimator performances, the opposite can occur around the maximum/minimum points and dependence on the data set size N arises, as a consequence of the higher statistical weight of data located in $x \simeq x_M$.

In the following simulations, the first problem, when x is large, is practically not meet, while the second problem emerges and the dependence on N is studied.

3.2. The choice of the kernel function K and the bandwidth h

The kernel function K can be chosen using the following physical arguments. The concentration field is related to the particle probability density function. In fact, let $N(\mathbf{x})$ and V be the number of particles located in $\mathbf{x}$ and the volume under consideration, respectively, then the concentration field is $c(\mathbf{x}) \propto N(\mathbf{x})/V$. However, $\langle N(\mathbf{x})\rangle \propto [\int p(\mathbf{x}|\mathbf{x}_0) S(\mathbf{x}_0) d\mathbf{x}_0] d\mathbf{x}$ (Section 2), finally the concentration field c is homogeneous if and only if $p(\mathbf{x}|\mathbf{x}_0)$ is constant all over the volume V. In (8) the spatial inhomogeneity is totally included in the kernel function K, then, from the above arguments, a physically sound choice of the kernel function is

$$K(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{x}_0) S(\mathbf{x}_0) d\mathbf{x}_0. \tag{12}$$

From (12) it emerges that, in the physical condition of uniformly distributed particles, a homogeneous concentration field is obtained.

Generally, h is chosen as an estimator performing parameter (de Haan, 1999; Vitali et al., 2006). However also an optimal bandwidth can be mathematically derived (Hall et al., 1994; Hall and Kang, 2005; Pope, 2000).

In fact, let Q be the function to be estimated and $\widehat{Q}$ its estimation. Following (Pope, 2000, §12.6.3–4), the error δ_Q in the kernel estimation is

$$\delta_Q = \widehat{Q} - Q, \tag{13}$$

which is composed by a deterministic part $B_Q = \langle \delta_Q \rangle$ called bias and a stochastic part with zero mean, unit variance and amplitude S_Q

$$\delta_Q = B_Q + S_Q \xi, \quad \langle \xi \rangle = 0, \quad \left\langle \xi^2 \right\rangle = 1. \tag{14}$$

The bias B_Q arises because the estimation is based on data that are not located in the point under consideration, the statistical error S_Q arises from the finite sample size, i.e. $N < \infty$. The mean-square error in the kernel estimation turns out to be

$$\left\langle \delta_Q^2 \right\rangle = B_Q^2 + S_Q^2. \tag{15}$$

In a d-dimensional case, it is possible to prove that (Pope, 2000; de Haan, 1999)

$$B_Q^2 \simeq \frac{1}{4}\theta^2 h^4 \left[\nabla^2 Q\right]^2 + \mathcal{O}\left(h^4\right), \tag{16}$$

and

$$S_Q^2 \simeq \frac{\gamma}{N h^d}, \tag{17}$$

where

$$\int x_i x_j K(\mathbf{x}, h) d\mathbf{x} = \theta h^2 \delta_{ij}, \quad \int K(\mathbf{x}, h)^2 d\mathbf{x} = \frac{\gamma}{h^d}. \tag{18}$$

Substituting (16) and (17) in (15) gives

$$\left\langle \delta_Q^2 \right\rangle \simeq \frac{1}{4}\theta^2 h^4 + \frac{\gamma}{N h^d}. \tag{19}$$

The optimal bandwidth is obtained from (19) setting its derivative with respect to h equals zero and it turns out to be determined as

$$h \simeq \left[d \frac{\gamma}{\theta^2}\right]^{1/(d+4)} N^{-1/(d+4)}. \tag{20}$$

Since in the following simulations one-dimensional case is considered, the size of the bandwidth is stated as

$$h \propto N^{-1/5}. \tag{21}$$

From (19) and (20) it is possible to observe that the quantity $\langle \delta_Q^2 \rangle^{1/2}$ goes to zero when N goes to infinite and, since h is a function of N (20) while K is not, it follows that the error in kernel estimation is mainly controlled by h than by K, which controls only the constant factors γ and θ (18). Moreover, using the idea of *equivalent bandwidths* (de Haan, 1999, §4.2), given the optimal bandwidth h_1 for a kernel K_1, it is possible to compute exactly the optimal bandwidth h_2 for a second kernel K_2:

$$\frac{h_1}{h_2} = \left[\frac{\gamma_1/\theta_1^2}{\gamma_2/\theta_2^2}\right]^{1/(d+4)}. \tag{22}$$

Finally, in general, nonparametric estimation is stronger dependent on the proper choice of the bandwidth h, than on the choice of the shape of the kernel function K, and it plays the role of a smoothing parameter (de Haan, 1999).

4. Numerical tests

4.1. Construction of an artificial dataset and estimator efficiency

In this section the estimator $\widehat{I}_S$ (9, 7–8) is numerically tested. Before to proceed an artificial dataset with known statistics must be constructed. To construct this dataset the function I_S computed by Galmarini et al. (1995) is considered as a reference. The authors plotted in their Fig. 4 the profile of I_S inside the cross-section of a Gaussian reactive plume dispersed into the atmospheric neutral boundary layer. The plot shows I_S as a function of the distance from the plume centreline r scaled on the plume standard deviation σ_r at different distances from the release point. Their function I_S is sufficiently well reproduced by

$$I_S = -A \frac{r^2}{\sigma_r^2} \exp\left\{-\frac{r^2}{\sigma_r^2}\right\}, \tag{23}$$

where in order to simulate four different distances from the source $A = 0.5, 1, 2, 4$ and the lower bound -1 is used when necessary, see Fig. 1. The values of A move from 4 to 0.5 when the distance from the source increases. A similar functional form of I_S has been given also in Rubio et al. (2008) to fit available data of

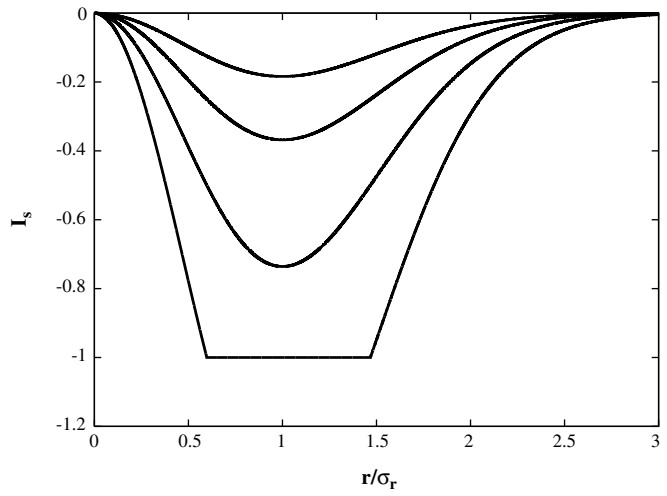

Fig. 1. The intensity of segregation functions I_S (23) with $A = 0.5, 1, 2, 4$ from the top to the bottom.

segregation of bimolecular reactive processes in porous media. The intensity of segregation is zero when $r = 0$ because solely the reactant brought by the plume is present and when $r/\sigma_r \gg 1$ because solely the environmental reactant is present. The minimum value of I_S occurs when $r/\sigma_r \simeq 1$, it equals -1 when measured near to the source and it increases when the distance from the source increases, because the reactants become more and more uniformly mixed. This well understandable situation is chosen as a reference paradigm to test the nonparametrical estimator proposed in Section 3.

Any sketch of the dispersion process, i.e. the cross-section of the plume at fixed distance from the source, is reproduced generating N random variables $\{r_i; i = 1, N\}$ from a Gaussian density with zero mean and variance $\langle r^2 \rangle = \sigma_r^2$. In order to study the dependence on the size of the dataset N, the simulations are performed with: $N = 100$, $N = 1000$ and $N = 50\,000$.

Let z_i be the position of the i-th particle and Z the plume centreline, then $z_i = r_i + Z$ and $\langle z^2 \rangle = \langle r^2 \rangle + Z^2$. Without loss in generality it can be stated $Z = 0$ and $\sigma_r = \sigma_z = 1$. For each z_i value a pair of correlated variable $\{c_\alpha(z_i),\ c_\beta(z_i)\}$ are randomly generated by a bivariate Gaussian density with given marginal mean concentrations $C_\alpha(z) = \langle c_\alpha(z) \rangle$, $C_\beta(z) = \langle c_\beta(z) \rangle$ and correlation of concentration fluctuations $R'_{\alpha\beta}(z) = \langle c'_\alpha(z) c'_\beta(z) \rangle$. The correlation function is defined as $R'_{\alpha\beta}(z) = \rho(z)\sigma_\alpha(z)\sigma_\beta(z)$ and formula (3) becomes

$$I_S = \rho \frac{\sigma_\alpha \sigma_\beta}{C_\alpha C_\beta}, \tag{24}$$

where ρ is the correlation coefficient, $\sigma_\alpha^2 = \langle c'^2_\alpha \rangle$ and $\sigma_\beta^2 = \langle c'^2_\beta \rangle$. For mathematical simplicity and random generation reasons, it is stated

$$\sigma_\alpha^2 = C_\beta^2 = \exp\left\{ -\frac{z^2}{\sigma_z^2} \right\} + 1,\ \sigma_\beta^2 = C_\alpha^2 = \exp\left\{ -\frac{z}{\sigma_z} \right\} + 1, \tag{25}$$

so that $\rho(z) = I_S(z)$. The kernel function is chosen Gaussian as the particle position density function

$$K\left(\frac{z - z_i}{h}\right) = \frac{1}{\sqrt{2\pi}h} \exp\left\{ -\frac{(z - z_i)^2}{2h^2} \right\}, \tag{26}$$

and the bandwidth parameter $h = N^{-1/5}\sigma_z$, in agreement with above discussion (Section 3.2), where the factor σ_z is for dimensional reasons. Finally, the physical bounds $-1 \le \widehat{I}_S \le 0$ are considered. In Fig. 2 it is shown the comparison between the exact function I_S (23) and the estimator $\widehat{I}_S$ for the three cases: $N = 100$, $N = 1000$ and $N = 50\,000$. The collective plots, the four curves I_S for the same values of N, are shown in Fig. 3. Generally, the estimation improves when the number of particles increases. For cases $N = 1000$ and $N = 50\,000$, the comparison is good and the estimator captures qualitatively the behaviour of the exact function. When $N = 1000$

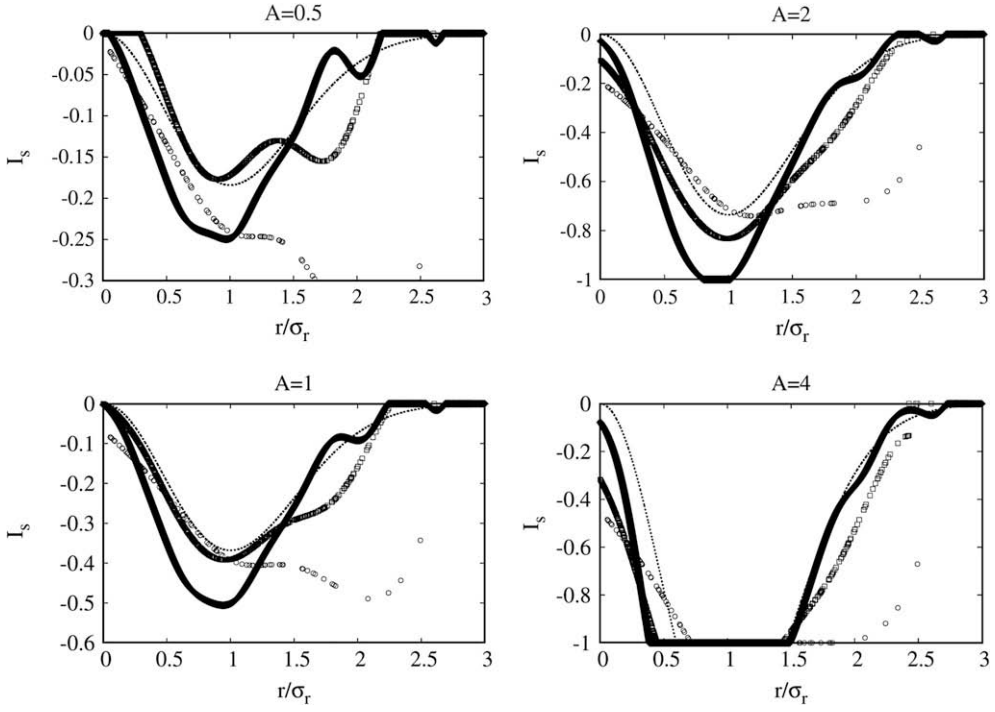

Fig. 2. The exact function I_S (dotted lines) and the estimator $\widehat{I}_S$ in the four numerical tests: $A = 0.5, 1, 2, 4$. Open circle $N = 100$, open square $N = 1000$, black diamond $N = 50{,}000$.

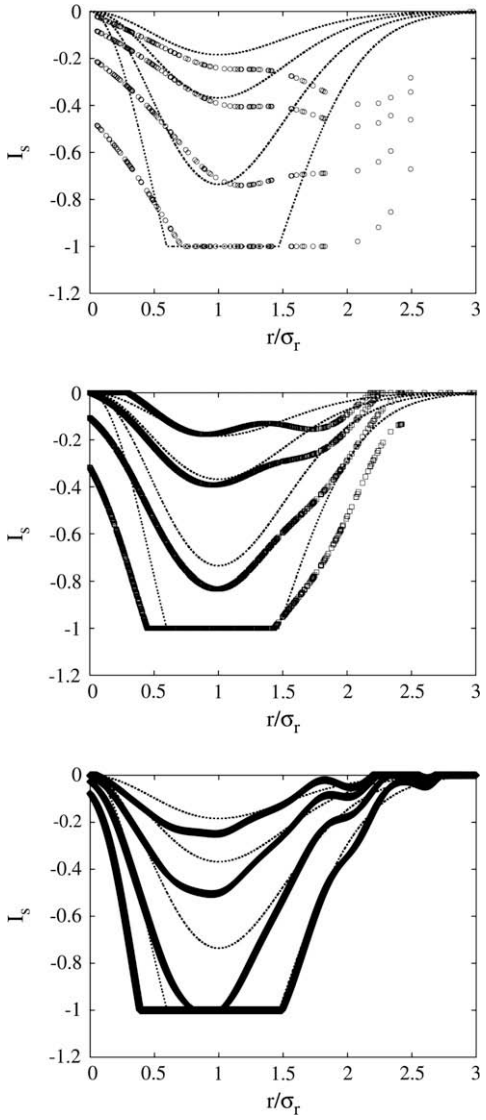

Fig. 3. Collective plot of the exact functions I_S (dotted lines) and the estimator $\widehat{I}_S$ in the four numerical tests: $A = 0.5, 1, 2, 4$. Open circle $N = 100$, open square $N = 1000$, black diamond $N = 50{,}000$.

the estimator works very well for small and moderate values of r/σ_r and becomes inaccurate for large values. However, looking plots for both $N = 1000$ and $N = 50\,000$, this less of accuracy is mainly due to the reduction of data for increasing r/σ_r than to the first remark in Section 3.1. In fact, in this case there are less data when r/σ_r increases as follows from the Gaussian particle distribution. Moreover, as remarked in Section 3.1, the dependency of the estimation on the sample size N around the minimum point is confirmed by the reduction of performance from $N = 1000$ to $N = 50\,000$ around $r/\sigma_r = 1$. More, when the exact function is steeply decreasing (or increasing), as cases $A = 2$ and $A = 4$, the estimator works less good in particular when $N = 1000$, because of the smoothness is driven by a bandwidth with a fixed functional form for all four cases. However this lack can be reduced using opportune sub-ensembles of the dataset, as discussed below in Section 4.2.

The main factor of chemical kinetic is the reaction rate k that for inhomogeneous mixing is replaced by an effective reaction rate $k_{eff} = k[1 + I_S]$ (9). To estimate the efficiency of the intensity of segregation estimator the percentage error of k_{eff} can be introduced

$$\Delta = \frac{k_{eff} - \widehat{k}_{eff}}{k_{eff}} = \frac{I_S - \widehat{I}_S}{1 + I_S}. \tag{27}$$

In Fig. 4 the percentage error (27) is plotted. It confirms that when N increases the accuracy increases. Moreover, when $r/\sigma_r \simeq 1$ the case $N = 50\,000$ has bad performances. In the first two cases $A = 0.5, 1$ the estimator with $N = 1000$ is surprisingly well working, generating a percentage error less than 5% for small and moderate values of r/σ_r and around 10% in the worst region, the same occurs, but with exchanged intervals, with $N = 50\,000$. The case $A = 2$ presents when $N = 1000$ an increasing percentage error that, however, remains bounded under 40%, while when $N = 50\,000$ a high percentage error occurs for small r/σ_r (up to 100%) and a very low error $\leq$5% for large r/σ_r. In the last case, $A = 4$, the percentage error is $\leq$40% and $\leq$10% for large r/σ_r when $N = 1000$ and $N = 50\,000$, respectively. Except for a steep increasing region around $r/\sigma_r \simeq 0.5$ where a 100% error occurs and a perfect estimation interval $\Delta = 0$ around $r/\sigma_r \simeq 1$. These last two particular results are just mathematical consequences of the physical limit $\{I_S, \widehat{I}_S\} \geq -1$. In fact, when $r/\sigma_r \simeq 0.5$ it is $\widehat{I}_S = -1$ and then $\Delta = [I_S - (-1)]/(1 + I_S) = 1$, and when $r/\sigma_r \simeq 1$ it is $I_S = \widehat{I}_S = -1$ for construction. Finally, the significance of this two extreme results is few and an efficiency similar to the previous case $A = 2$ can be considered.

After the above analysis, what concern the performance dependence on the sample size N, it is possible to conclude that the ensemble data $N = 1000$ is sufficiently good performing taking into account also the computational effort required by the case $N = 50\,000$.

4.2. Sub-ensemble statistics

In this section a method to improve the estimator performance is shown. This method consists by using a sub-ensemble of the dataset, instead of the whole, for each point that belongs to a critical interval. This sub-ensemble includes only the data that are more closed to the point under consideration. It follows that for each point a different dataset is used.

Looking at Figs. 2–4, one observes that with $N = 1000$, the performances in the interval $0 \leq r/\sigma_r \leq 0.5$ when $A = 2$ and $A = 4$ are reduced with respect to those when $A = 0.5$ and $A = 1$. Since the curve to be estimated in these cases is steeply decreasing, the reduction of efficiency is due to the data located in distant points. In other words, to a high gradient follows a high value difference between points and then data from distant locations invalidate the estimation.

The proposed idea of sub-ensemble is to use only those data that are more closely located to the point under consideration. If the interval $0 \leq r/\sigma_r \leq 0.5$ is considered then for each r, as a case study, only the sub-ensemble of data r_i such that $|r - r_i|/h \leq 0.5$ is used to compute the estimation by (9), (7)–(8).

The results are shown in Fig. 5 where the previous estimation with the whole dataset and the new with the sub-ensemble data for the interval $0 \leq r/\sigma_r \leq 0.5$ are plotted together and compared with the exact function. The estimator based on the sub-ensemble data has a better performance in the interval under consideration. From this it follows that also the percentage error is reduced in the same range of r/σ_r, as is shown in Fig. 6.

Finally, an improvement of the estimator performance can be obtained using for each point of a critical interval only that sub-ensemble of data which are more closed to the considered point. This last result is a remarkable property of the kernel method that gives support to the present suggestion to use the nonparametric approach to measure the intensity of segregation of second-order chemical reactions in Lagrangian air quality modelling.

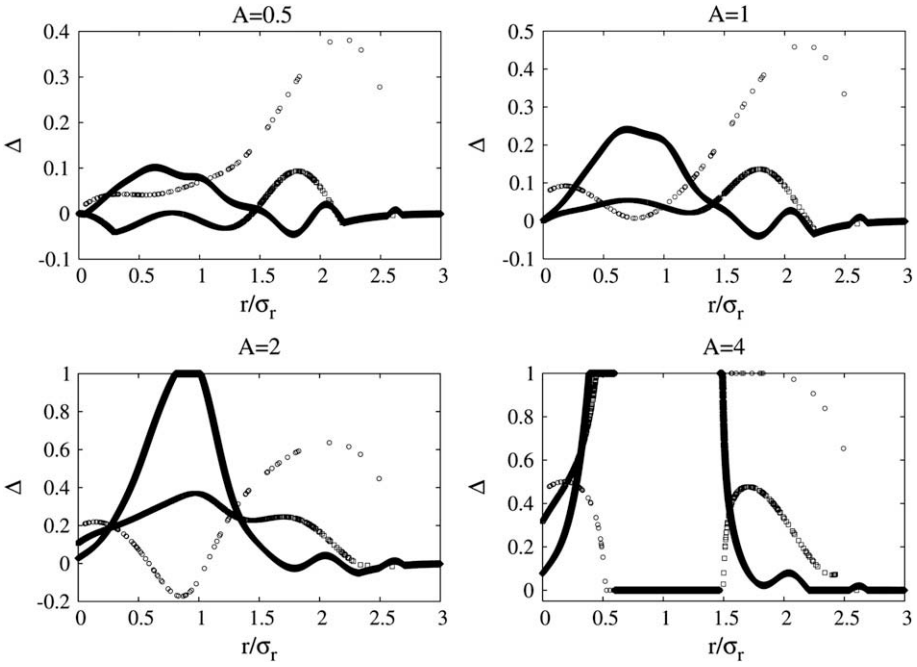

Fig. 4. Plot of the percentage error Δ in the four numerical tests: $A = 0.5, 1, 2, 4$. Open circle $N = 100$, open square $N = 1000$, black diamond $N = 50{,}000$.

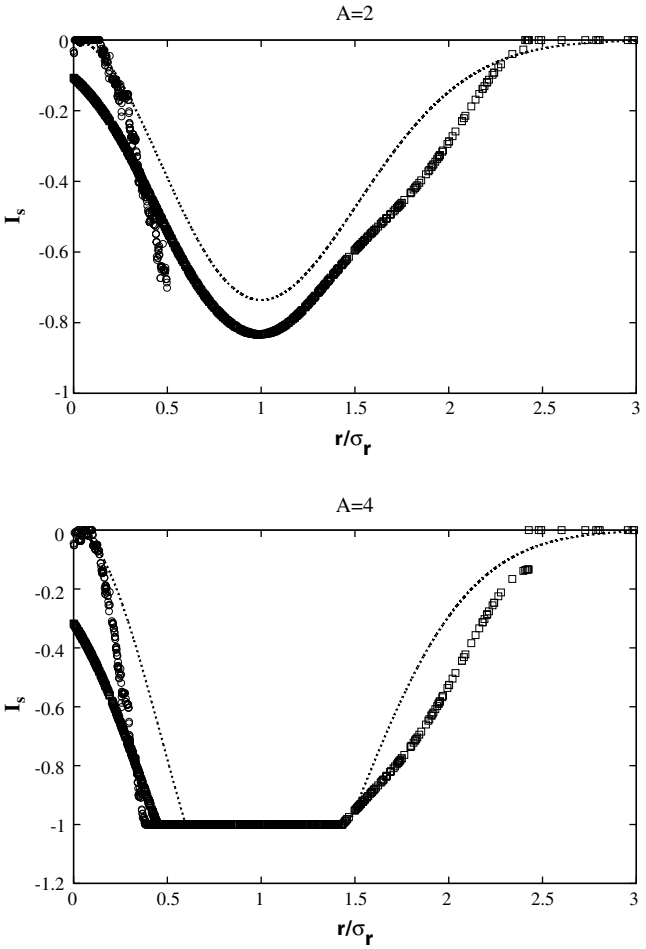

Fig. 5. The exact functions I_S (dotted lines) and the estimator $\widehat{I}_S$ with the whole dataset □ and with a sub-ensemble ○.

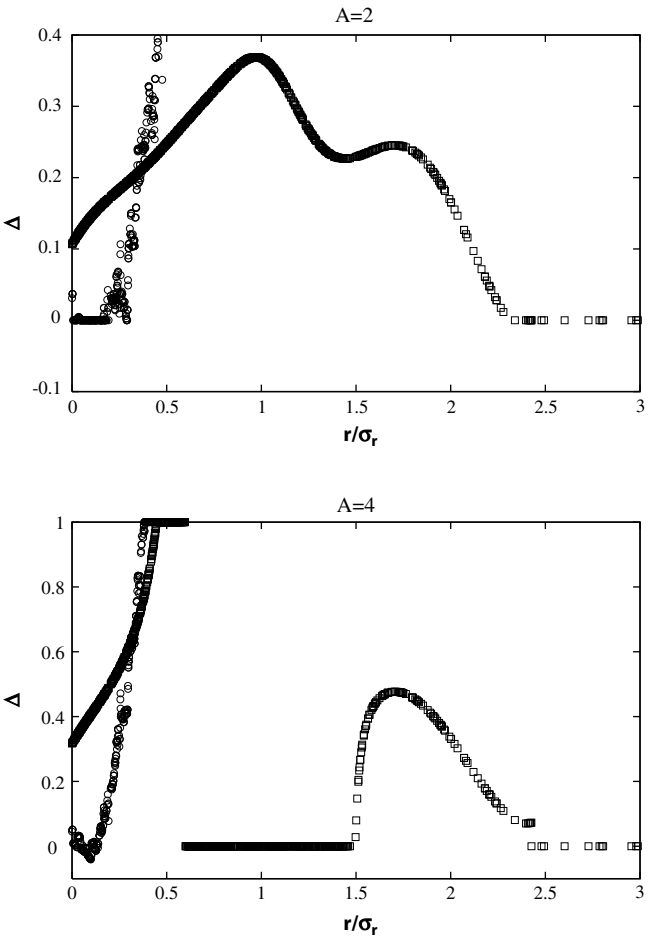

Fig. 6. Plot of the percentage error Δ for the four intensity of segregation estimators with the whole dataset □ and with a sub-ensemble ○.

5. Discussion and conclusions

In the present paper the kernel method is used in order to propose a nonparametric estimator of the intensity of segregation in second-order chemical reactions and its efficiency is explored. Since the estimator proposed is grid-free it is particularly suitable for those grid-free Lagrangian air quality models that, using the kernel method, have been recently formulated for concentration field computation. The probability density of the positions of the particles bringing the reactants is chosen as kernel function. This fact permits the estimator to take into account inhomogeneity on a physical ground. A remarkable feature of this approach is that a statistical quantity generated by an inhomogeneous ensemble, such as the intensity of segregation, is computed along the realization of the flow. This is really important for the evolution in time of chemical transformations during the running of each simulation of the dispersion process.

Numerical tests are performed with an artificial dataset generated to reproduce the intensity of segregation function inside the cross-section of a reactive plume with Gaussian distribution profile, at different distances from the source, as computed by Galmarini et al. (1995) in a neutral atmospheric boundary layer. The estimator has been tested for three sample sizes: $N = 100$, $N = 1000$ and $N = 50\,000$. Generally the performance improves when N increases. However, as remarked in Section 3.1, a dependence on N can arise around the maximum/minimum points. This fact has been observed in simulations. Balancing accuracy and computational costs, the dataset with $N = 1000$ is the preferable one. The estimator works well especially for small and moderate separations from the plume centreline and generally in smooth intervals of the function. The effective reaction rate is computed and the percentage error emerges to be less than 5% in the best estimation ranges and less than 40% in the worst. A method to reduce percentage error in the worst cases is introduced. It consists in using the sub-ensemble of data which includes only those data that are more closely located to the point under consideration. As a consequence, for each point a different dataset is used. When this method is applied to estimate the function in points inside a critical interval, the performance improves.

These results support the approach proposed here and its applications in Lagrangian air quality modelling.

Acknowledgments

The author is grateful to the Editor for helpfulness and kindness and to the anonymous Referees for their criticism and suggestions that improved the manuscript. Furthermore, G. Zanini, M. Mazzeo, F. Monforti, L. Vitali and G. Pace at ENEA Bologna are acknowledged for support, Prof. Yehua Li for useful references on nonparametric estimation and Prof. Kee-Hoon Kang to have made available to me the extended and more detailed version of his paper (Hall and Kang, 2005). The research has been financed by Regione Emilia-Romagna in the frame of the project "Laboratorio LaRIA".

References

Aguirre, C., Brizuela, A., Vinkovic, I., Simoëns, S., 2005. A subgrid Lagrangian stochastic models for turbulent passive and reactive scalar dispersion. Int. J. Heat Fluid Flow 27, 627–635.

de Arellano, J.V.-G., Dosio, A., Vinuesa, J.-F., Holtslag, A., Galmarini, S., 2004. The dispersion of chemically reactive species in the atmospheric boundary layer. Meteor. Atmos. Phys. 87, 23–38.

de Arellano, J.V.-G., 2003. Bridging the gap between atmospheric physics and chemistry in studies of small-scale turbulence. Bull. Amer. Meteor. Soc. 84 (1), 51–56.

Chock, D., Winkler, S., Sun, P., 2002. Effect of grid resolution and subgrid assumptions on the model prediction of a reactive bouyant plume under convective conditions. Atmos. Environ. 36, 4649–4662.

Cokca, E., 2003. A computer program for the analysis of 1-D contaminant migration through a soil layer. Environ. Model. & Software 18, 147–153.

Corrsin, S., 1958. Statistical behavior of a reacting mixture in isotropic turbulence. Phys. Fluids 1 (1), 42–47.

Crone, G., Dinar, N., van Dop, H., Verver, G., 1999. A Lagrangian approach for modelling turbulent transport and chemistry. Atmos. Environ. 33, 4919–4934.

Danckwerts, P., 1952. The definition and the measurement of some characteristics of mixtures. Appl. Sci. Res. A 3, 279–296.

Davakis, E., Nychas, S., Andronopoulos, S., 2003. Validation study of the dispersion Lagrangian particle model DIPCOT over complex topographies using different concentration calculation methods. Int. J. Environ. Pollut. 20, 33–46.

Donaldson, C. duP., Hilst, G., 1972. Effects of inhomogeneous mixing on atmospheric photochemical reactions. Environ. Sci. Technol. 6 (9), 812–816.

van Dop, H., 2001. The evaluation of a Lagrangian model for turbulent transport and chemistry. Phys. Fluids 13 (5), 1331–1342.

Dopazo, C., Valino, L., Fueyo, N., 1997. Statistical description of the turbulent mixing of scalar fields. Int. J. Mod. Phys. B 11 (25), 2975–3014.

Galmarini, S., de Arellano, J.V.-G., Duynkerke, P., 1995. The effect of micro-scale turbulence on the reaction rate in a chemically reactive plume. Atmos. Environ. 29 (1), 87–95.

Garmory, A., Richardson, E., Mastorakos, E., 2006. Micromixing effects in a reactive plume by the Stochastic Fields Methods. Atmos. Environ. 40, 1078–1091.

de Haan, P., 1999. On the use of density kernels for concentration estimations within particle and puff dispersion model. Atmos. Environ. 33, 2007–2021.

Hall, P., Fisher, N., Hoffmann, B., 1994. On the nonparametric estimation of covariance functions. Ann. Statist. 22 (4), 2115–2134.

Hall, P., Kang, K.-H., 2005. Bandwidth choice for nonparametrical classification. Ann. Statist. 33 (1), 284–306. Extended version available at http://stats.hufs.ac.kr/%7Ekhkang.

Hall, P., Patil, P., 1994. Properties of nonparametric estimators of autocovariance for stationary random fields. Probab. Theory Relat. Fields 99, 399–424.

Hill, J., 1976. Homogeneous turbulent mixing with chemical reaction. Ann. Rev. Fluid Mech. 8, 135–161.

Hilst, G., 1998. Segregation and chemical reaction rates in air quality models. Atmos. Environ. 22, 3891–3895.

Karamchandani, P., Santos, L., Sykes, I., Zhang, Y., Tonne, C., Seigneur, C., 2000. Development and evaluation of a State-of-the-Science reactive plume model. Environ. Sci. Technol. 34, 870–880.

Klimenko, A.Y., Bilger, R.W., 1999. Conditional moment closure for turbulent combustion. Prog. Energy Combust. Sci. 25, 595–687.

Komori, S., Hunt, J., Kanzaki, T., Murakami, Y., 1991. The effects of turbulent mixing on the correlation between two species and on concentrations in non-premixed reacting flows. J. Fluid Mech. 228, 629–659.

Krol, M., Molemaker, M., de Arellano, J.V.-G., 2000. Effects of turbulence and heterogeneous emissions on photochemically active species in the convective boundary layer. J. Geophys. Res. 105 (D5), 6871–6884.

Lamb, R., 1973. Note on the application of the K-theory to diffusion problems involving nonlinear chemical reactions. Atmos. Environ. 7, 257–263.

Lamb, R., Shu, W., 1978. A model of second-order chemical reactions in turbulent fluid – part I. Formulation and validation. Atmos. Environ. 12, 1685–1694.

Lorimer, G., 1986. The kernel method for air quality modelling – I. Mathematical foundation. Atmos. Environ. 20, 1447–1452.

Lorimer, G., Ross, D., 1986. The kernel method for air quality modelling – II. Comparison with analytic solutions. Atmos. Environ. 20, 1773–1780.

Meeder, J., Nieuwstadt, F., 2000. Large-eddy simulation of turbulent dispersion of a reactive plume from point source into a neutral atmospheric boundary layer. Atmos. Environ. 34, 3563–3573.

Molemaker, M., de Arellano, J.V.-G., 1998. Control of chemical reactions by convective turbulence in the boundary layer. J. Atmos. Sci. 55, 568–579.

Monforti, F., Vitali, L., Pagnini, G., Lorenzini, R., Monache, L.D., Zanini, G., 2006. Testing kernel density reconstruction for Lagrangian photochemical modelling. Atmos. Environ. 40, 7770–7785.

Petersen, A., Holtslag, A., 1999. A first-order closure for covariances and fluxes of reactive species in the convective boundary layer. J. Appl. Meteor. 38, 1758–1776.

Pope, S., 1985. PDF methods for turbulent reactive flows. Prog. Energy Combust. Sci. 11, 119–192.

Pope, S., 2000. Turbulent Flows. Cambridge University Press.

Rubio, A., Zalts, A., Hasi, C.E., 2008. Numerical solution of the advection–reaction–diffusion equation at different scales. Environ. Model. & Software 23, 90–95.

Schumann, U., 1989. Large-eddy simulation of turbulent diffusion with chemical reactions in the convective boundary layer. Atmos. Environ. 23, 1713–1727.

Shu, W., Lamb, R., Seinfeld, J., 1978. A model of second-order chemical reactions in turbulent fluid – part II. Application to atmospheric plumes. Atmos. Environ. 12, 1695–1704.

Stockwell, W., 1995. Effects of turbulence on gas-phase atmospheric chemistry: calculation of the relationship between time scales for diffusion and chemical reaction. Meteorol. Atmos. Phys. 57, 159–171.

Sykes, R., Henn, D., Parker, S., Lewellen, W., 1992. Large-eddy simulation of turbulent reacting plume. Atmos. Environ. 26A, 2565–2574.

Sykes, R., Parker, S., Henn, D., Lewellen, W., 1994. Turbulent mixing with chemical reaction in the planetary boundary layer. J. Appl. Meteor. 33, 825–834.

Valiño, L., 1998. A field Monte Carlo formulation for calculating the probability density function of a single scalar in a turbulent flow. Flow, Turb. Combustion 60, 157–172.

Vinuesa, J.-F., de Arellano, J.V.-G., 2003. Fluxes and (co-)variances of reacting scalars in the convective boundary layer. Tellus 55B, 935–949.

Vinuesa, J.-F., de Arellano, J.V.-G., 2005. Introducing effective reaction rates to account for the inefficient mixing of the convective boundary layer. Atmos. Environ. 39, 445–461.

Vitali, L., Monforti, F., Bellasio, R., Bianconi, R., Sacchero, V., Mosca, S., Zanini, G., 2006. Validation of a Lagrangian dispersion model implementing different kernel methods for density reconstruction. Atmos. Environ. 40, 8020–8033.