



Improving Numerical Dispersion Modelling in Built Environments with Data Assimilation Using the Iterative Ensemble Kalman Smoother

Cécile L. Defforge¹ · Bertrand Carissimo¹  · Marc Bocquet¹ · Raphaël Bresson¹ · Patrick Armand²

Received: 24 September 2019 / Accepted: 4 November 2020 / Published online: 14 January 2021
© The Author(s) 2021

Abstract

Air-pollution modelling at the local scale requires accurate meteorological inputs such as from the velocity field. These meteorological fields are generally simulated with microscale models (here *Code_Saturne*), which are forced with boundary conditions provided by larger scale models or observations. Local atmospheric simulations are very sensitive to the boundary conditions, whose accurate estimation is difficult but crucial. When observations of the wind speed and turbulence or pollutant concentration are available inside the domain, they provide supplementary information via data assimilation, to enhance the simulation accuracy by modifying the boundary conditions. Among the existing data assimilation methods, the iterative ensemble Kalman smoother (IEnKS) is adapted to urban-scale simulations. This method has already been found to increase the accuracy of wind-resource assessment. Here we assess the ability of the IEnKS method to improve scalar-dispersion modelling—an important component of air-quality modelling—by assimilating perturbed measurements inside the urban canopy. To test the data assimilation method in urban conditions, we use the observations provided by the Mock Urban Setting Test field campaign and consider cases with neutral and stable conditions, and the boundary conditions consisting of the horizontal velocity components and turbulence. We prove the capacity of the IEnKS method to assimilate observations of velocity as well as pollutant concentration. In both cases, the accuracy of pollutant concentration estimates is enhanced by 40–60%. We also show that assimilating both types of observations allows further improvements of turbulence predictions by the model.

Keywords Data assimilation · Iterative ensemble · Kalman smoother · Micrometeorology · Mock Urban Setting Test · Urban air quality

✉ Bertrand Carissimo
bertrand.carissimo@enpc.fr

¹ CEREAs, Joint laboratory, École des Ponts ParisTech and EDF R&D, 6 Quai Watier, 78400 Chatou, France

² CEA, DAM, DIF, 91297 Arpajon, France

1 Introduction

Atmospheric-dispersion modelling is used for numerous applications: air-quality assessment, exposure indicators, supporting air-quality planning and directives, or for nuclear emergencies, such as an accident or malicious activity (e.g., Hanna et al. 2006; Kumar et al. 2011; Benamrane et al. 2013). Many of these applications occur at the local scale in urban areas or around industrial sites where the presence of buildings makes the simulation domain more complex. Since the evolution of pollutants in the atmosphere is mainly driven by the flow, dispersion modelling requires knowledge of the meteorology over small domains.

Computational fluid dynamics models, together with the Reynolds-averaged Navier–Stokes model for turbulence, include only a few approximations and are well suited to precisely account for buildings and to simulate the flow field and temperature over complex terrain. As a consequence, there is a growing consensus for using computational uid dynamics (CFD) models in atmospheric-dispersion modelling at the local scale (Robins 2003; Holmes and Morawska 2006; Blocken et al. 2013). Such models generally offer the possibility of transporting passive scalars and thus representing the evolution of non-reactive tracer concentration. In order to represent reactive pollutants, CFD models can be coupled to chemistry or aerosol models (Albriet et al. 2010).

Computational fluid dynamics models used for dispersion modelling are very sensitive to the input parameters related to the pollutant source term (location and mass rate) and to the meteorological conditions. In particular, the domains under scrutiny in micrometeorological studies are relatively small compared with the distance travelled by flow information during a typical integration time. Finally, the influence of the initial conditions rapidly vanishes and the solution is largely determined by the boundary conditions which have a substantial impact on the accuracy of the local-scale atmospheric simulations (Srebric et al. 2008).

Observations are widely used for dispersion studies but mainly for validation purposes. Currently, the exploitation of observations through data assimilation (DA) has mostly been performed to solve inverse problems for source identification (e.g., Davoine and Bocquet 2007; Winiarek et al. 2012). A few studies have used DA methods to improve the knowledge of the source and flow field input parameters for puff models (Krysta et al. 2006; Kovalets et al. 2009). In such studies, the assimilated observations are measurements of pollutant concentration and the typical spatial scales are much larger than those considered in micrometeorology. So far, the flow observations available within built environments have rarely been included to improve micrometeorological simulations. While these observations might be strongly perturbed by the geometrical features of the domain, they still contain useful information which could be taken into account through the DA method.

In the DA literature, boundary conditions are rarely considered as control variables. Only recently, Mons et al. (2017) and Sousa et al. (2018) have used ensemble-based DA methods to estimate the incoming wind speed and direction at one level for input to the CFD simulations, including only these two variables in the control vector. In the present work, we propose describing the boundary conditions with a complete vertical profile of velocity (u - and v -components) and turbulence kinetic energy (TKE), which leads to tens of control variables. We thus use the DA method to enhance the accuracy of CFD atmospheric simulations at the local scale through the correction of a large control vector of the boundary conditions.

Among the existing DA methods, the iterative ensemble Kalman smoother (IEnKS) was chosen and adapted to local-scale simulations by taking the boundary conditions into account. The IEnKS method was selected after comparing it with the back-and-forth nudging algorithm (Auroux and Blum 2008; Defforge et al. 2018); both approaches present the advantage of

being independent of the model, in particular because they do not require the adjoint model. In addition, the IEnKS method is well adapted to parallel computation and to nonlinear systems such as the Navier–Stokes equations. Previous studies and alternative methods have already proved the ability of the IEnKS method to assimilate flow observations available within prospected sites for the installation of wind farms, which has improved the accuracy of wind-resource assessment (Defforge et al. 2019). Since we consider stationary boundary conditions and assume that the model is perfect, there is no difference between the filtering (IEnKF, Sakov et al. 2012) and smoothing (IEnKS) formulations. Anticipating future applications with unsteady boundary conditions, we consider the smoothing method in what follows, keeping in mind that it is here strictly equivalent to the filtering method.

Below, we evaluate the ability of the IEnKS method, in urban conditions, to assimilate in situ velocity and concentration observations, perturbed by obstacles, to correct the meteorological boundary conditions prescribed for atmospheric simulations with a CFD model. To achieve this goal, we use the measurements provided by the Mock Urban Setting Test (MUST) campaign, described in Sect. 2. In Sect. 3, we briefly present the IEnKS algorithm, followed by the experimental set-up and the methods developed to apply the IEnKS method to this study case. Section 4 summarizes the results obtained with the IEnKS method by assimilating first only the flow observations, then only the concentration observations, and finally both types of observations together. Our conclusions and some perspectives to further improve the performance of the IEnKS method are given in Sect. 5.

2 The Mock Urban Setting Test Campaign

The MUST campaign was conducted in September 2001 at the U.S. Army Dugway Proving Ground test site (Biltoft 2001; Yee and Biltoft 2004). The aim of this field campaign was to provide well-documented cases of urban dispersion by acquiring both meteorological and dispersion observations. These datasets have been widely used for urban-model development and validation (e.g. Milliez and Carissimo 2007, 2008; Winiarek 2014; Bahlali et al. 2019).

The site is located in the Utah desert where containers form a regular array over a $200 \times 200 \text{ m}^2$ area. The containers are aligned on a 12 by 10 grid and each container is 12.2 m long, 2.42 m wide, and 2.54 m high. Several trials were performed, with each consisting of a 15-min release of a tracer gas (propylene) from a fixed position, either within or immediately outside the container array, at a height between 0.15 and 5.2 m. Biltoft (2001) and Yee and Biltoft (2004) give a comprehensive description of the experimental set-up and analyses of the results. Numerous instruments were provided by several organisations, such as the Army Research Laboratory (ARL), Arizona State University (ASU), the Dugway Proving Ground (DPG), the Defence Science Technology Laboratory, the Los Alamos National Laboratory (LANL), and the the University of Utah (UU).

2.1 Meteorological Measurements

Many meteorological instruments were installed within and outside the MUST array during the release experiments. Figure 1 shows the MUST domain with the container array and the meteorological instruments within the domain. In the experiments, all the containers representing the buildings are not exactly identical, and the position of the instruments is approximately known. As with previous authors, and with these reasons in mind, we also have idealized the grid with an identical aligned container to simplify the mesh.

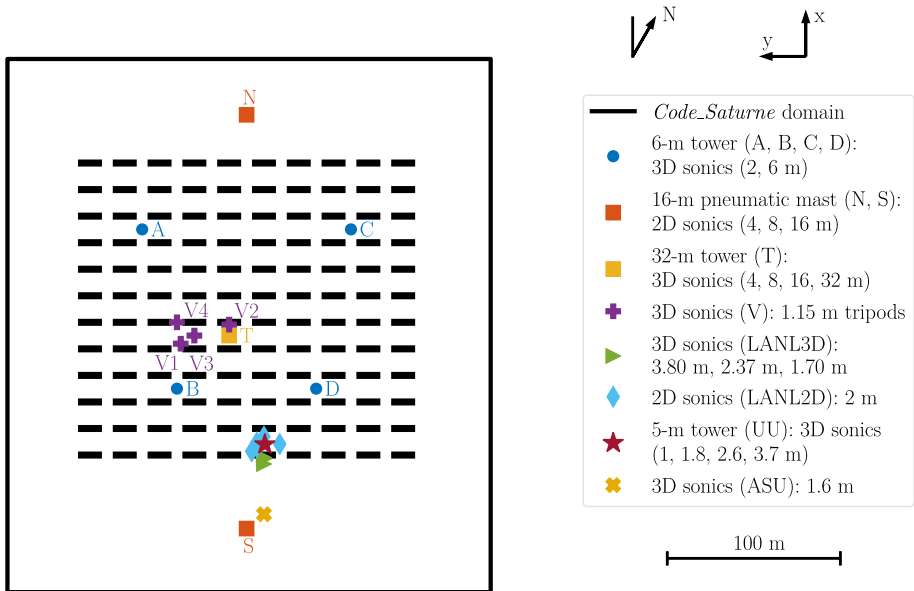


Fig. 1 Representation of the MUST array within the domain used for the *Code_Saturne* simulations; adapted from Biltoft (2001). The location of all meteorological instruments (such as the three- and two-dimensional (3D, 2D) sonic anemometers) is also shown

Within the domain, numerous two- and three-dimensional (2D, 3D) sonic anemometers established to provide observations of the velocity at a frequency of 10 Hz. Information about these instruments is given in the legend of Fig. 1 and more details are available in Biltoft (2001). As in Milliez (2006), for the 2D sonic anemometers we estimate the 3D TKE from its 2D value using

$$k_{3D} = 1.3k_{2D}. \quad (1)$$

which is approximate and valid for the neutral and stable conditions considered here; moreover, the 2D sonic anemometers are not used in the assimilation but only the validation.

For each 15-min trial, Yee and Biltoft (2004) extracted one 200-s quasi-steady period during which the meteorological variables as well as gas concentration are nearly constant. We aim to represent the velocity and concentration fields during two of these periods as stationary simulations with a CFD model with a Reynolds-averaged Navier–Stokes turbulence closure. To be consistent with this closure, the observations assimilated correspond to the mean values, averaged over the 200-s period (i.e. 2000 sonic observations).

Other instruments are permanently installed in the vicinity of the MUST array and provide supplementary information. We use observations provided by the Portable Weather Information and Display System (PWIDS) and the Surface Atmospheric Measurement System (SAMS), located less than 2 km upstream of the containers array. A sound detection and ranging (sodar) system, located 400 m north-east of the domain, also provides flow measurements, especially in upper layers of the atmospheric boundary layer (see Fig. 21). These instruments provide observations of the horizontal velocity components (u , v) at 2 m above the ground for the PWIDS station, at 10 m for the the SAMS station, and at several levels between 15 and 200 m above ground level in 5-m intervals for the sodar.

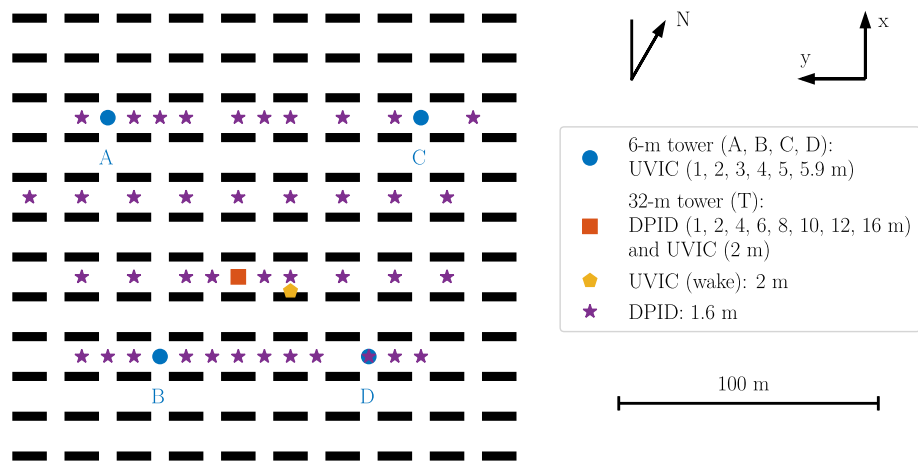


Fig. 2 Representation of the MUST array with the location of the instruments measuring gas concentration, adapted from Biltoft (2001)

2.2 Gas-Concentration Measurements

The instruments for measuring gas concentration installed in the domain are of two types:

- ultraviolet ion collector (UVIC) installed on the 6-m towers, near the 32-m tower and at the centre of the array. These instruments were calibrated over a range of 0.01–1000 ppmv.
- digital photo-ionization detectors (digiPID) located on the 32-m tower and along four east–west lines inside the containers array with a calibrated operating range of 0.04–1000 ppmv.

Figure 2 shows the location of the 48 digiPID and 26 UVIC instruments. The observations of gas concentration are available during all the 15-min trials with a frequency of 50 Hz such that the assimilated values are averaged over 10,000 observations (200 s).

3 Methods

3.1 Experimental Set-Up

The meteorological fields in the region containing the MUST array are simulated using the atmospheric module of *Code_Saturne* (Archambeau et al. 2004), which is developed by the Centre d'Enseignement et de Recherche en Environnement Atmosphérique (CEREA) to simulate atmospheric flows in the atmospheric boundary layer. This numerical model has been used to simulate 20 different releases during the MUST campaign with varying meteorological conditions (see Milliez and Carissimo 2007 for details). Here we mention that, in the basic CFD code, an option has been introduced for atmospheric simulations, which includes the equation of energy written with the potential temperature variable, modifications of the buoyancy production/destruction of turbulence, and a modification for the stratified wind-speed profile over a rough surface (Monin–Obukhov similarity theory).

The domain used for the simulations is $348 \times 348 \text{ m}^2$ horizontally and extends up to 50 m vertically. The domain is centred on the container array and includes a band of flat terrain

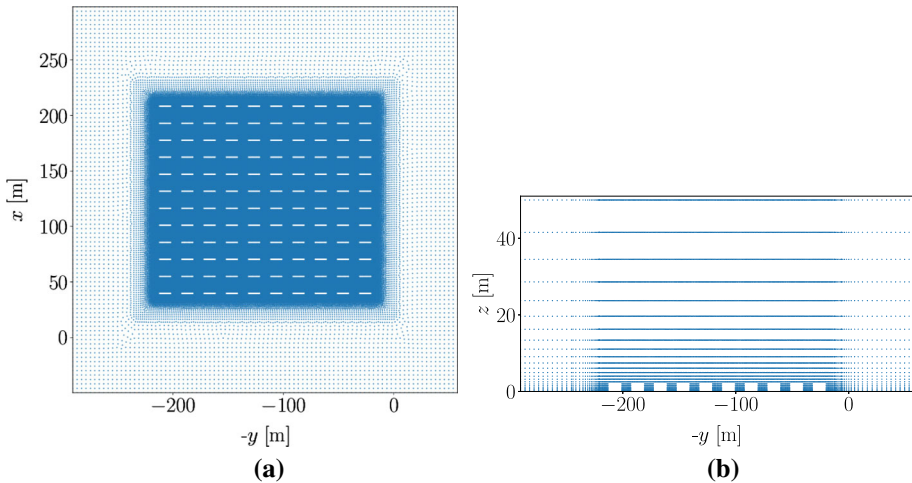


Fig. 3 Mesh used for the *Code_Saturne* simulations. The points represent the cell edges. **(a)** Cross-section at constant $h = 1.5$ m and **(b)** a cross-section at a constant $x = 115.5$ m

outside the array (77 m on the eastern and western sides and 87 m on the northern and southern sides). The horizontal resolution is somewhat coarse outside the MUST array ($\Delta x = \Delta y = 4$ m) and refined inside the container array down to 0.5 m. The vertical resolution decreases with altitude from $\Delta z = 0.25$ m near the ground to $\Delta z = 7.8$ m for the uppermost layer (see Fig. 3), and the containers representing the buildings were not exactly identical. In addition, the positions of the instruments were only approximately known.

The turbulence is modelled using the $k - \epsilon$ model, where k is the TKE and ϵ its dissipation rate. Since we use a Reynolds-averaged Navier–Stokes closure model, all the simulated variables correspond to ensemble means.

In our method, we can impose several profiles on the boundary to represent spatial gradients, which are very small here and so we used only one profile arbitrarily located at the centre of the face, implying the same boundary conditions over all the faces. Therefore, the boundary conditions included in the control vector correspond to one vertical profile, arbitrarily located in the middle of the southern border of the domain. From this vertical profile of boundary conditions prescribed as model input, the *Code_Saturne* model extrapolates the values to all the boundary faces. The boundary conditions are considered constant in all the studies presented here. The profile is defined by 22 vertical levels for the three variables: u , v , and k . At the first timestep of the model, the field of the dissipation rate is computed in order to ensure the equilibrium between turbulence production and dissipation. The values calculated at the boundaries of the domain are kept as boundary conditions of the dissipation rate for the next integration steps. This process is performed for each simulation to adapt the ϵ boundary conditions to the prescribed boundary conditions. The model is integrated for 5000 iterations with a fixed timestep of 0.1 s, which is enough to reach a steady state (Bahlali 2018). Note that over the IEnKS algorithm iterations, the boundary conditions recalculated and prescribed to the model may not be in equilibrium with the model physics. As a result, the flow may evolve between the border boundaries, where the boundary conditions are prescribed, and the first obstacles. We verified that this evolution is small enough.

We consider two different trials of the MUST campaign:

- The neutral case (2681829) which was recorded on the 25 September 2011 at 1829 LT (local time = UTC –6h). The meteorological conditions at this time corresponded to neutral stability with an estimated Obukhov length of $L = 28,000$ m (Yee and Bilotft 2004). The pollutant source is located between two containers in the bottom-left corner of the domain, at $h = 1.8$ m above the ground.
- The stable case (2692157) which was recorded on the 26 September 2011 at 2157 LT. The conditions were stable with an estimated Obukhov length of $L = 130$ m (Yee and Bilotft 2004). The tracer gas is released from the roof of the first container in the bottom-right corner, at $h = 2.6$ m above the ground.

3.2 The Iterative Ensemble Kalman Smoother

The IEnKS algorithm (Bocquet and Sakov 2014) is an ensemble variational DA method. As a variational method, it is based on the minimization of a cost function and as an ensemble-based method, the analysis error space is spanned by a limited number of vectors consisting of the ensemble members. We consider stationary boundary conditions, referred to as vector \mathbf{z} of size l , for which a first estimate, or background, is available: \mathbf{z}^b . The background-error statistics are represented by an ensemble of N vectors of boundary conditions, centred on \mathbf{z}^b : \mathbf{E}^b . The matrix \mathbf{E}^b is of size $l \times N$ such that the ensemble members correspond to its columns. We can thus define the (normalized) anomalies as the departure of each member from the background as

$$\mathbf{A} = \frac{1}{\sqrt{N-1}} [\mathbf{E}^b - \mathbf{z}^b \mathbf{1}^T], \tag{2}$$

where $\mathbf{1}$ is a vector of size N with all components equal to one. The available observations are referred to as the vector \mathbf{y} of size p and the observation-error covariance matrix is \mathbf{R} .

The IEnKS cost function derived for micrometeorological applications (Defforge et al. 2018, 2019) is the same as that obtained in previous DA schemes (e.g., Bocquet and Sakov 2014), albeit replacing the initial conditions by the boundary conditions

$$\tilde{\mathcal{J}}(\mathbf{w}) = \frac{1}{2} \|\mathbf{y} - \mathcal{F}(\mathbf{z}^b + \mathbf{A}\mathbf{w})\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \|\mathbf{w}\|^2, \tag{3}$$

where \mathbf{w} is the weight vector of size N and \mathcal{F} is the forward operator, which returns the simulated observations for a given vector of the boundary conditions, and we use the notation $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$ and $\|\mathbf{x}\|_{\mathbf{Y}}^2 = \mathbf{x}^T \mathbf{Y} \mathbf{x}$. In practice, this operator corresponds to the integration of the *Code_Saturne* model forced with the boundary conditions and the extraction of the simulated values at the position of the available observations. The variables at the end of the analysis are referred to with the superscript “a”.

Here, the cost function (3) is minimized following the Gauss–Newton algorithm. Alternatively, we could use other minimization schemes such as the Levenberg–Marquardt approach (Björck Å 1996). A significant advantage of working in the ensemble space is that the calculation of the gradient of the cost function requires neither the full (in state space) adjoint nor the tangent linear of the forward operator \mathcal{F} . One analysis cycle is schematically explained in Fig. 4.

Background ensemble: $\mathbf{E}^b = \underbrace{\mathbf{z}^b \mathbf{1}^T}_{\text{mean}} + \underbrace{\sqrt{N-1} \mathbf{A}}_{\text{anomalies}}$ Initialization: $\mathbf{w} = 0$

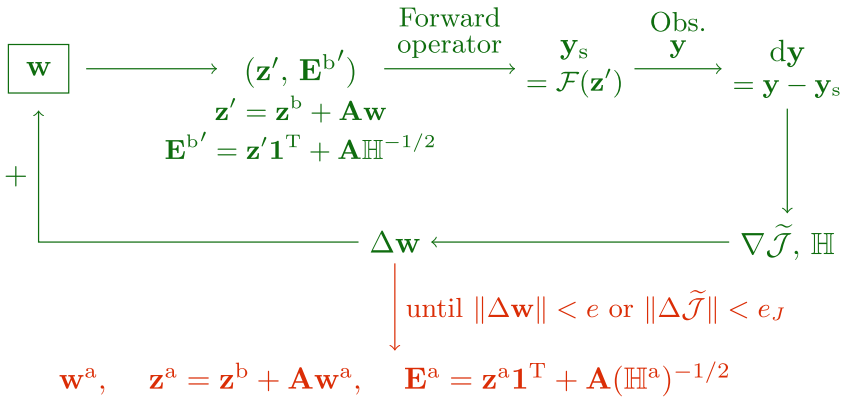


Fig. 4 One analysis cycle of the IEnKS algorithm. The background ensemble \mathbf{E}^b is an input of the method. The best estimate of the weight vector \mathbf{w}^a is obtained by minimizing the cost function $\tilde{\mathcal{J}}$ as shown by the cycle: for each value of \mathbf{w} , a new ensemble of BC ($\mathbf{E}^{b'}$), centred on \mathbf{z}' , is transformed using the ensemble transform matrix ($\mathbf{T} = \mathbb{H}^{-1/2}$) obtained at the previous iteration. The forward operator is applied to this transformed ensemble, which gives an ensemble of simulated observations (Obs.) with mean \mathbf{y}_s that can be compared to the observations \mathbf{y} . The increment $d\mathbf{y} = \mathbf{y} - \mathbf{y}_s$ and the spread of the ensemble of simulated observations around the mean are used in the estimation of the gradient and approximate Hessian (i.e., without the second-order derivative correction) of the cost function ($\nabla \tilde{\mathcal{J}}$ and \mathbb{H}). The weight vector \mathbf{w} is thus updated following the Gauss–Newton algorithm until the convergence criterion is reached, defined as thresholds for the change in \mathbf{w} (e) or in $\tilde{\mathcal{J}}$ (e_J). At the end of the analysis, the best estimate of the control vector \mathbf{z}^a and the analysis ensemble \mathbf{E}^a can be used as a first guess for the next analysis cycle in general cases. In the particular case studied here (steady state), only one analysis cycle is performed

3.3 Anamorphosis for the Turbulence Kinetic Energy

In the DA experiment presented here, the control vectors correspond to the micrometeorological boundary conditions for the flow field. While the background errors relative to the horizontal velocity components u and v can be considered as Gaussian variables (not shown), it is certainly not a valid hypothesis for the TKE, which is always positive. Analyzing observations of the TKE during all the trials shows that its probability density function approximately follows an exponential distribution (see Fig. 20b in the Appendix).

The cumulative distribution function (c.d.f.) of the exponential distribution is

$$F(k) = 1 - \exp(-\lambda k), \tag{4}$$

where λ is the rate parameter. Since the IEnKS method assumes a Gaussian behavior of the control variables, we must use anamorphosis for the TKE.

The principle of anamorphosis is to find a bijection between the c.d.f. of a non-Gaussian variable (here k) and a Gaussian variable which is included in the control vector in place of the non-Gaussian variable (Cohn 1997; Bertino et al. 2003). We thus define a new, non-physical variable γ following the normal law $n(0, \sigma_\gamma)$ and its c.d.f. $G(\gamma)$, such that

$$k = F^{-1} \circ G(\gamma) \tag{5}$$

$$\gamma = G^{-1} \circ F(k). \tag{6}$$

The choice of the standard deviation σ_γ of the parameter γ does not affect the DA experiment in general. Here, as we apply an ensemble-based DA technique, the construction of the ensemble is clarified in the next section.

The control vector considered includes the 22 values of the variables u , v , and γ and its total size is $l = 66$. The forward operator \mathcal{F} includes one more step to convert the values of γ into TKE values, then prescribed as boundary conditions for the TKE.

3.4 Construction of the Background Ensemble

In order to construct the background ensemble necessary for the DA experiment with the IEnKS method, we first estimate the background-error covariance matrix, \mathbf{B} . The ensemble anomalies are then determined as the leading modes of this matrix such that

$$\mathbf{B} = \mathbf{A}\mathbf{A}^T. \tag{7}$$

The coefficients of the background-error covariance matrix are decomposed as the product of correlation coefficients $\mathbf{C}_{i,j}$ and standard deviation of the individual background errors σ_i ,

$$\mathbf{B}_{i,j} = \mathbf{C}_{i,j}\sigma_i\sigma_j. \tag{8}$$

In order to determine the coefficients $\mathbf{C}_{i,j}$ and σ_i , we analyzed the measurements available for all the trials within the domain and above the containers, choosing to describe the vertical correlations with a Balgovind function (Balgovind et al. 1983) which is usually applied to represent the spatial structure of error statistics (Winiarek 2014). Based on the climatology, we set the correlation length $R = 3$ m (see Appendix 1 for more details).

Using the climatological analyses and expert judgement, we set the background error standard deviation for the velocity components u and v to $\sigma_{uv}^b = 5 \text{ m s}^{-1}$. For the TKE, we estimated the rate parameter to be $\lambda_k = 1.25 \text{ s}^2 \text{ m}^{-2}$ and we set the standard deviation of the anamorphosis variable γ to $\sigma_\gamma^b = 5$ to give as much weight to the velocity components as TKE in the DA experiment.

Eventually, we obtain an estimate of the background-error covariance matrix used to define the background ensemble. The background ensemble corresponds to the $N - 1$ leading modes of the background-error covariance matrix—i.e., eigenvectors associated with the largest eigenvalues, with a N th member necessary to recentre the ensemble. The ensemble of N members thus formed is rotated using a random rotation matrix to obtain a recentred ensemble with statistically equivalent members.

3.5 Assimilating Observations of Concentration

Values of gas concentration are always positive, such that the distribution of the observation errors cannot be considered as Gaussian. This issue can be overcome by assuming that the observation error for concentration observations follows a log-normal distribution (see Liu et al. 2017). Liu et al. also highlighted that the logarithmic function may give too much weight to small concentration values, which is why they proposed a threshold-constant value c_t in the logarithmic function; the innovation vector thus becomes

$$\mathbf{dy} = \ln(c_t + \mathbf{y}) - \ln(c_t + \mathcal{F}(\mathbf{z})). \tag{9}$$

In the DA experiments involving concentration observations, we set $c_t = 0.05$ ppmv.

3.6 Observation-Error Covariance Matrix

The observation-error covariance matrix \mathbf{R} is diagonal, and the coefficients on the diagonal are equal to

$$(\sigma^o)^2 N_{xy}, \quad (10)$$

where $(\sigma^o)^2$ is the variance associated with the observation error and N_{xy} is the number of observations available, and assimilated, at the same horizontal position but different heights. This scaling factor aims at giving similar weight in the cost function to the different areas of the domain and to not favour regions that are more observed (more details are given in Appendix 2).

For flow observations, the standard deviation of the observation error is set to $\sigma_{uv}^o = 0.2 \text{ m s}^{-1}$, and for the concentration observations (after applying the logarithmic transformation defined by Eq. 9, we set $\sigma_c^o = 0.05$.

3.7 Estimation of the Background Boundary Conditions

The background boundary conditions are estimated from observations provided by instruments outside of the container array, referred to as “background observations” (see Table 1). The profile of the wind speed is assumed to follow a semi-logarithmic law given by Monin–Obukhov similarity theory (Stull 1988). The surface stress is estimated from the background observations: $u_* = 0.57 \text{ m s}^{-1}$ in the neutral case and $u_* = 0.40 \text{ m s}^{-1}$ in the stable case. The profile of TKE is also obtained from the Monin–Obukhov similarity theory.

The vertical profile of the wind direction is obtained as the sum of a constant value, the average of the background observations, and an empirical perturbation, which is equal to 10° near the ground and vanishes with height. In the neutral case, the mean wind direction is equal to -42.2° , and in the stable case it is estimated as 45.4° . Note that these angles correspond to the flow direction, measured anticlockwise from the x -axis. More details about the construction of the background boundary conditions can be found in Appendix 3.

3.8 Reference Simulations

The MUST campaign has already been used in previous studies for validation purposes (Milliez 2006; Winiarek 2014; Bahlali 2018). The CFD simulations performed in these studies corresponded to the 200-s quasi-steady periods for each 15-min experiment, selected by Yee and Bilstoft (2004). The boundary conditions were estimated from the observations provided by the sonic anemometers installed on the 16-m tower at the southern edge of the domain (see Fig. 21). Since the tower is very close to the location where the boundary conditions are prescribed, the simulation obtained with these boundary conditions are very close to the observations within the domain. Comparing the wind speeds in these previous simulations to all the available flow observations, we calculate that the mean error is 0.5 m s^{-1} in both the neutral and stable cases. Consequently, the simulations performed in these previous studies are considered as reference simulations. The results obtained here with the IEnKS method, using different input data, are compared with these reference simulations.

4 Results with the Iterative Ensemble Kalman Smoother and Field Measurements

4.1 Experimental Set-Up

As previously detailed, the domain used for the numerical simulations with the CFD model *Code_Saturne* is of size $348\text{ m} \times 348\text{ m} \times 50\text{ m}$ with unstructured mesh. The control vector corresponds to a vertical profile of boundary conditions located in the middle of the southern boundary of the domain, and defined by 22 vertical levels for three variables: u , v , and γ (which is the anamorphosis variable representing k), giving a control vector of size $l = 66$.

The neutral and stable cases were treated slightly differently and the differences of parametrization are summarized in Table 1.

For both neutral and stable cases, we show the results of three DA experiments aiming at correcting the boundary conditions related to the parameters u , v , and k .

1. The first experiment, referred to as “W” for wind only, consists of assimilating 14 observations of the two components of the horizontal velocity (u , v) with the IEnKS algorithm.
2. In the second experiment, referred to as “C” for concentration only, the IEnKS algorithm is applied to assimilate 13 observations of tracer gas concentration.
3. The last experiment, referred to as “WC” for wind and concentration, considers observations of both (u , v) and c . In order to keep a similar number of observations in total, we considered 15 observations: four of u , four of v , and seven of concentration.

The selected measurements for these three experiments in the neutral and ntable cases are detailed in Table 1.

As mentioned in Sect. 3.4, the background-error covariance matrix has diagonal values $(\sigma_{uv}^b)^2 = 25\text{ m}^2\text{ s}^{-2}$ and the observation-error covariance matrix is diagonal with $\sigma_{uv}^o = 0.2\text{ m s}^{-1}$ for the u and v observations and $\sigma_c^o = 0.05$ for the c observations (see Sect. 3.6). These choices yield more weight to the observations than the background, which is representative of the operational conditions.

For all the experiments, the ensemble considered is composed of $N = 5$ members and the convergence criterion is set to $e_J = 0.01$, which means that the iterative minimization algorithm is stopped when the variation of the cost function between two iterations is smaller than 1% of the initial value of the cost function. If this criterion is not reached after $j_{\max} = 10$ iterations, the algorithm is stopped anyway. The analysis is set equal to the value of the control vector which leads to the minimum of the cost function within the available iterations—10 or less if the algorithm has converged before.

Cross-validation is performed with all the flow and concentration observations available during the studied trials and not assimilated in any of the three experiments. These observations are referred to as validation observations below. For the neutral case, there are 24 validation observations for the u velocity component, 24 for the v velocity component, 29 for the TKE, and 40 for the concentration. For the stable case, there are 24 validation observations for each of u and v components, 30 for the TKE, and 44 for the concentration. The small difference comes from instruments that were not operational during the neutral trial for unknown reasons.

Below, we present the results obtained for the two cases studied here (neutral and stable) and for each of the three different experiments.

Table 1 The background boundary conditions are estimated by fitting a semi-logarithmic velocity profile to the observations (obs.) provided by the instruments indicated in the first row (see Sect. 3.7)

| | Neutral | Stable |
|--|--|--|
| Background observations | PWIDS (2 m), SAMS (10 m), and sodar (45 m) | SAMS (10 m) and sodar (20 and 45 m) |
| θ perturbation | $\Delta\theta_0 = -10^\circ$ | $\Delta\theta_0 = -10^\circ$ |
| Experiment W: 14 obs. of (u, v) | Tower A at (2, 6 m), Sonic V3 (1.15 m), 5-m tower (1, 1.8, 2.6, 3.7 m) | Tower C (2, 6 m), Sonic V3 (1.15 m), 5-m tower (1, 1.8, 2.6, 3.7 m) |
| Experiment C: 13 obs. of c | Towers C (1, 2, 3 m) and D (1, 2, 3 m), 32-m tower (1, 2, 4, 6, 8, 10 m), DPID #26 (1.6 m) | Towers A (1, 2, 3 m) and B (1, 2, 3 m), 32-m tower (1, 2, 4, 6, 8, 10 m), DPID #23 (1.6 m) |
| Experiment WC: 15 obs. (8 of (u, v) and 7 of c) | (u, v): 5-m tower (1, 1.8, 2.6 m) and Sonic V3 (1.15 m), c : Towers C (1, 2 m), D (1, 2 m), and 32-m tower (1, 2, 4 m) | (u, v): 5-m tower (1, 1.8, 2.6 m) and Sonic V3 (1.15 m), c : Towers A (1, 2 m), B (1, 2 m), and 32-m tower (1, 2, 4 m) |

The vertical profile of wind direction (θ) is perturbed following Eq. 23 with a perturbation near the ground given by the second row (see Fig. 22c). The observations used in the three assimilation experiments (W, C, and WC) are indicated in the three last rows

4.2 Neutral Case

With the background boundary conditions constructed as explained in Appendix 3, the departure of the simulated velocity field from the available u and v observations selected for the DA method (see Table 1) is on average 1.44 m s^{-1} . With the parametrization given in Table 1, the optimal boundary conditions are obtained after eight IEnKS iterations in the experiment W, four in experiment C, and nine in experiment WC.

Figure 5a, b shows, for the reference simulation, the horizontal velocity field (arrows) and concentration (colours) at two constant heights above the ground: $h = 4 \text{ m}$ and $h = 1 \text{ m}$ respectively. Note that, for the sake of clarity, arrows are depicted every 15 m in Fig. 5a and every 2 m in Fig. 5b. For similar reasons, we represent the concentration field on a logarithmic scale in Fig. 5a, b. The departure from the reference simulation is computed for the background and the analyses of the three experiments. The error fields are shown at $h = 4 \text{ m}$ in Figs. 6a, 7a, 8a, and 9a, and at $h = 1 \text{ m}$, magnifying the vicinity of the pollutant source, in Figs. 6b, 7b, 8b, and 9b.

Figure 6a indicates that the background error for the velocity (arrows) is mostly along the y axis and in opposite direction compared with the mean flow, which means that the background velocity field underestimates the magnitude of the v -velocity component and the u -component slightly. As a result, the incidence angle is larger than the reference simulation (Fig. 5a) and the flow is less aligned with the containers. Moreover, the wind speed is slightly smaller in the background simulation than in the reference one. Consequently, the gas is less diluted by the airflow and the concentration in the plume is higher, which is shown by the positive values of concentration error larger than negative ones (see the colourscale in Fig. 6a). In addition, the background pollutant plume, which is well aligned with the wind direction at this height (above obstacle top), drifts more to the left than the reference plume. Consequently, the concentrations are overestimated to the left of the plume and underestimated on the right.

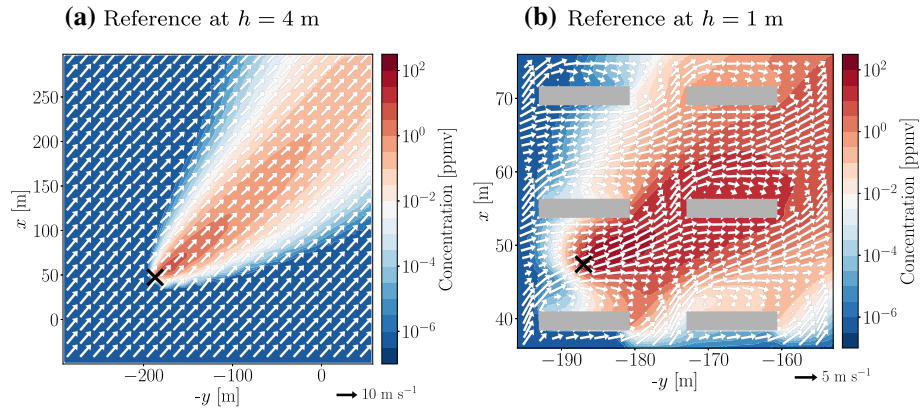


Fig. 5 Reference simulation for the neutral case. Horizontal cross-section of the concentration field in logarithmic scale (colours) and the velocity field (arrows) at **(a)** $h = 4$ m and **(b)** $h = 1$ m (zoom near the source). The black cross locates the source

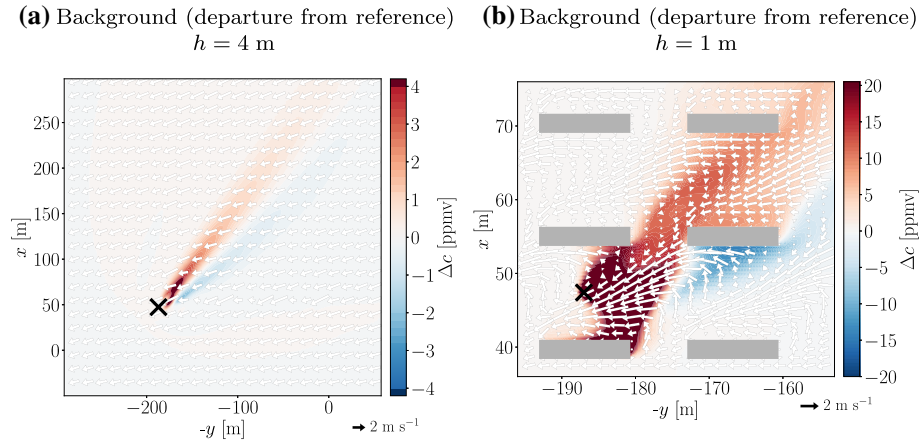


Fig. 6 Departure of the background simulation from the reference in the neutral case. The error field of concentration field (colours) and velocity (arrows) are plotted at **(a)** $h = 4$ m and **(b)** $h = 1$ m (zoom near the pollutant source). The black cross locates the source

Figure 5b shows that, below the canopy top, the presence of obstacles tends to slow and tilt the flow to the right, aligning it to the container array. Figure 6b shows the departure from this reference simulation in the background simulation. The larger departure between the background incidence angle and the alignment of the containers causes a greater effect of the obstacles on the background flow: the wind speed decelerates more within the container array in the background simulation than in the reference one. Moreover, we can observe large background errors in between the obstacles, indicating that the deviation of the flow due to the obstacles is stronger, which results in recirculation near the source. The variation of the deflection angle of the pollutant plume—within the container array—with the incidence angle, has been previously observed by Yee and Bilotft (2004). In the bottom-left region, the gas is propagated towards the south in the background simulation, whereas this effect is weaker in

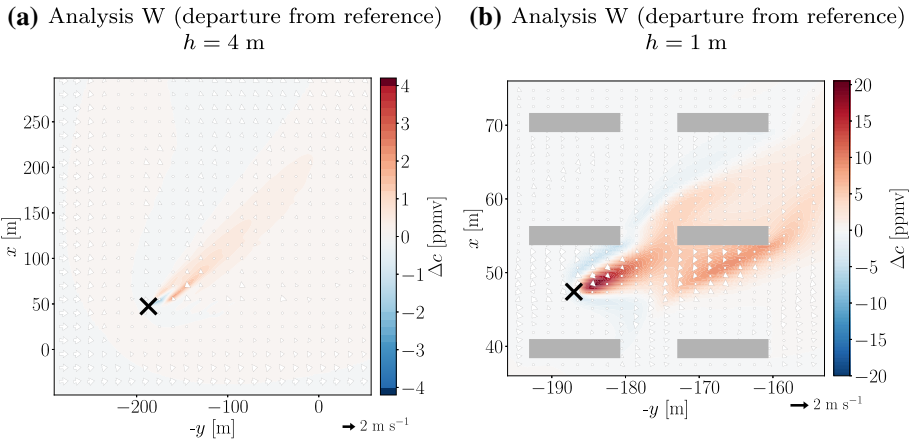


Fig. 7 As Fig. 6 but for the analysis state of the experiment W

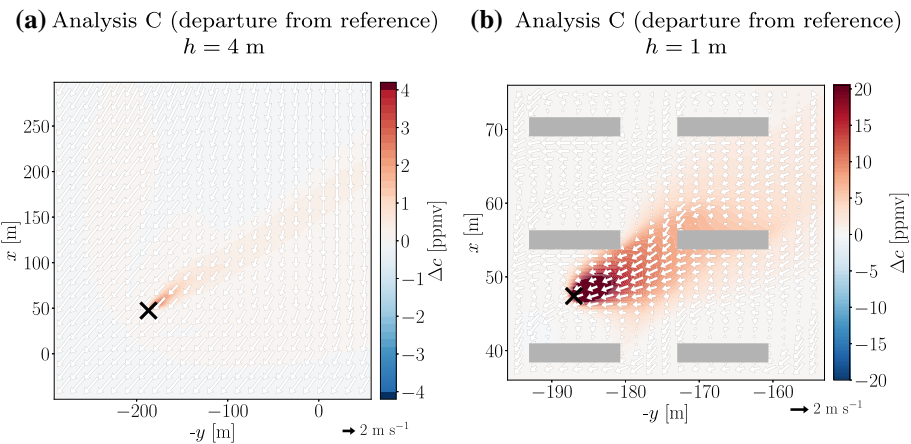


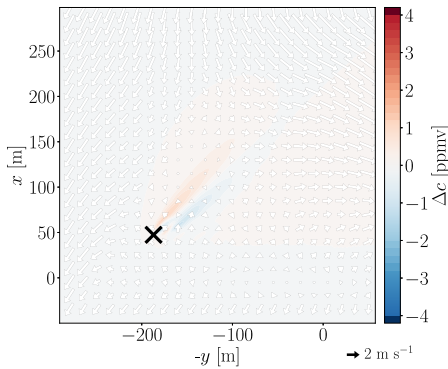
Fig. 8 As Fig. 6 but for the analysis state of the experiment C

the reference simulation. As a consequence, the background simulation substantially overestimates the concentrations in this region, as shown in Fig. 6b.

Figure 7a shows that in the analysis of experiment W, the error for the velocity field is largely reduced as compared with the background simulation, resulting in the pollutant plume more similar to the reference, with the error for the concentration field also reduced to very small values. Figure 7b shows the error fields for concentration and velocity within the urban canopy for the analysis of experiment W, illustrating the very good agreement with the reference in terms of wind direction, wind speed, and concentration. As a result, the analysis errors are substantially reduced as compared with the background simulation, but a slight overestimation of the gas concentration still persists.

Figure 8a, b shows that for the analysis of experiment C, the picture is more nuanced. Upstream, the error of velocity, shown in Fig. 8a, is somewhat aligned with the mean flow, which means that the wind direction has been corrected in the boundary conditions; however, the wind speed is still underestimated. Downstream, the downwards arrows indicate that the

(a) Analysis WC (departure from reference)
at $h = 4$ m



(b) Analysis WC at $h = 1$ m

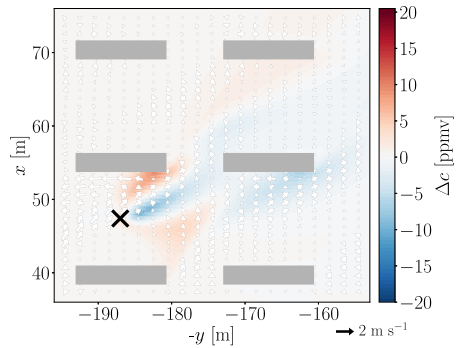


Fig. 9 As Fig. 6 but for the analysis state of the experiment WC

error for the v velocity component has been reduced but the error for the u velocity component has increased. As a result, the incidence angle of the flow is now too small and the pollutant plume far from the source goes more to the right than the reference simulation. The remaining errors for the wind speed and direction, or alternatively for the u and v velocity components, explain the slight overestimation of the concentrations on the right side of the plume. Still, the errors for concentration values are largely reduced between the background and analysis C. Figure 8b shows that, near the pollutant source, the wind speed is still underestimated, though the effect of the buildings is here less pronounced due to the better alignment of the flow with the containers. The concentrations are slightly overestimated in analysis C, though only along the gas plume axis (compare with Fig. 5b), which indicates that the shape of the pollutant plume simulated with the analysis C is very similar to the reference within the canopy. In particular, the concentrations decrease with the distance from the source at a similar rate and, in the vicinity of the source, the recirculation has vanished and there is no more overestimation of the concentrations south of the gas source.

Figure 9a, b shows the results of the experiment WC for which both the velocity and concentration observations are assimilated. The global picture of the error field (Fig. 9a) shows that the wind direction, and the velocity to a lesser extent, are corrected upstream of the containers, leading to very small errors within the container array, and resulting in the large reduction in the errors for the concentration field. The magnified view at $h = 1$ m in Fig. 9b shows errors that are localized and with small amplitudes, which means that both velocity and concentration fields simulated with the analysis WC are similar to the reference simulation. The results obtained in this experiment are noticeably better than those of experiments W and C.

In order to further quantify the benefit of the DA process, Fig. 10 shows the simulated versus observed values for the wind speed and direction, the TKE, and tracer concentration. These scatter plots are presented line by line, respectively, for the reference simulation, the background, and analyses of all three experiments. Selected error statistics—MAE, the r.m.s. error, and the Pearson coefficient—are also shown.

The error statistics shown in Fig. 10 are calculated by taking into account all the observations—even the assimilated ones—and without any specific averaging. However, some masts provide several observations at the same horizontal location but at different ver-

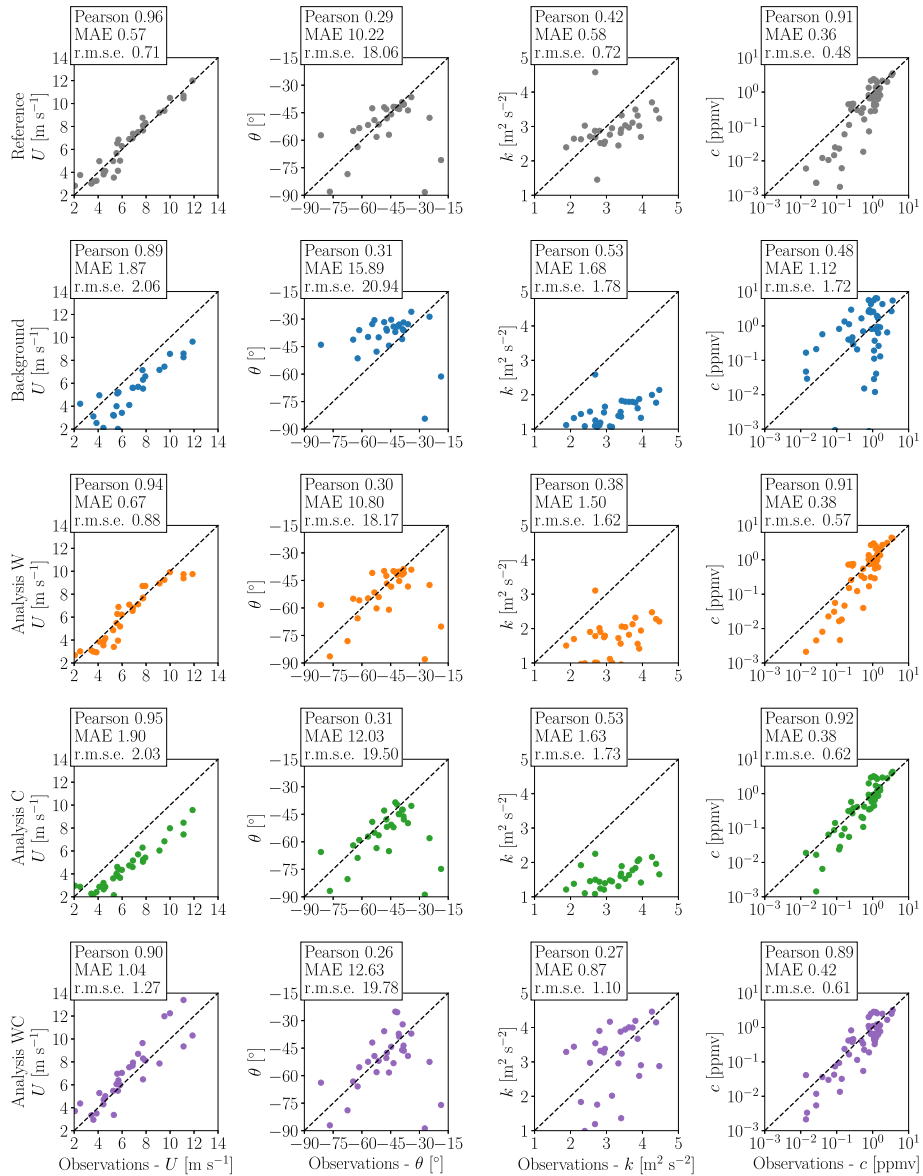


Fig. 10 Simulated versus observed values for the reference simulation (first row), the background (second row), Analysis W (third row), Analysis C (fourth row), and Analysis WC (last row) for the neutral case. The scatter plots are shown for wind speed and direction (U , θ), the TKE (k), and the concentration of the tracer gas (c). The Pearson correlation coefficient, the mean absolute error (MAE), and the r.m.s. error are also calculated for each variable and simulation. All the available observations are plotted and taken into account in the statistics calculations

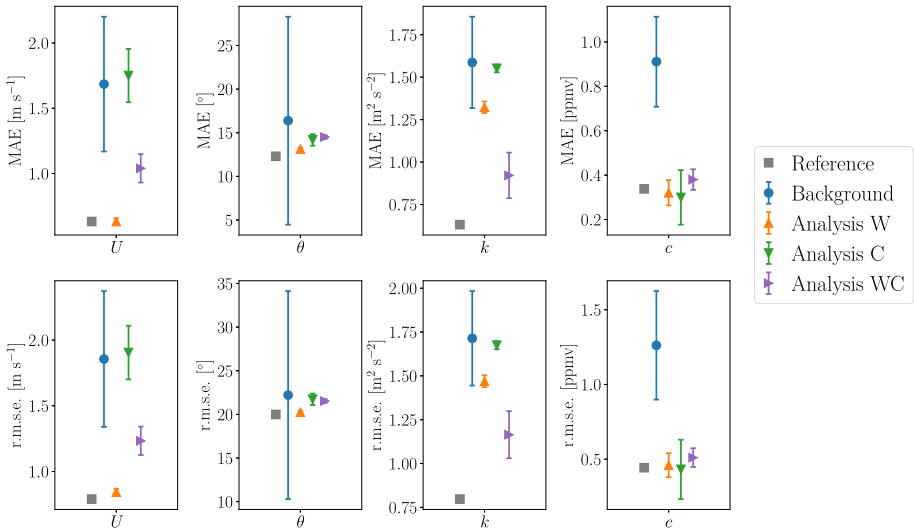


Fig. 11 Mean absolute error and r.m.s. error calculated over the validation observations of the wind speed and direction, k , and c within the urban canopy ($h < 2.5$ m) for the reference, background, and analyses W, C, and WC in the neutral case. The mean values are calculated as an average over the errors, weighted by the inverse of the number of observations available at the same location. Error bars represent twice the standard deviation of the background and analysis ensembles

tical levels. In order to avoid giving more weight to these densely-observed locations, the errors are multiplied by a coefficient equal to the inverse of the number of available observations (for the same variable) at the same horizontal position. The statistics shown in Fig. 11 are obtained with this weighting convention and considering only the validation observations; Fig. 11 also shows the standard deviation of the MAE and the r.m.s. error, calculated over the background and analysis ensembles (error bars).

From the quantitative statistics in Figs. 10 and 11, we confirm the conclusions drawn from the previous qualitative analysis of the velocity and concentration fields (Figs. 7, 8, 9). In the experiment W, the assimilation of velocity observations enables the reduction in most of the error for the velocity and direction to levels very close to that obtained with the reference simulation. As a result of a better estimation of the velocity field, there is a better agreement between the observed and simulated values of concentration and the error for the concentration is significantly reduced.

In contrast, assimilating observations of concentration does not necessarily improve the knowledge of the velocity field. In the present example, the error for the wind speed in analysis C is quite similar to that observed in the background (systematic underestimation), while the wind direction is significantly better estimated after assimilation. This result is consistent with the fact that dispersion—and especially the pattern of pollutant concentration—is particularly sensitive to the wind direction. Consequently, the errors of concentration estimations are largely reduced for the Analysis C. Dispersion, and thus the values of concentration, are the result of two additive effects: dilution, which increases with the wind speed and turbulent diffusion, which increases with the value of k . We can suppose that observations of concentration only are not enough for the DA algorithm to discriminate which of the two variables (wind speed or TKE) should be corrected. We can see in Fig. 10 that both wind

speed and TKE are underestimated here, which is consistent with the slight overestimation of concentrations near the source observed in Fig. 8b.

In all the simulations corresponding to the background, W and C analyses, the turbulence is underestimated and the IEnKS method does not help much to reduce this error. The small impact of the IEnKS method on turbulence boundary conditions (not shown) and consequently on the field of simulated TKE suggests that, in the present case, the inflow turbulence does not impact much the gas dispersion. In fact, most of the turbulence is formed due to the presence of obstacles. The results of analysis WC are quite similar to analysis W for the wind speed and direction, and improved for the variables TKE and c . In addition, the assimilation of both types of observations also enables the correction of the inflow turbulence, probably because the combined information from the velocity and concentration observations allows the discrimination of dilution and diffusion.

Figure 11 also shows the standard deviation of the updated ensembles, which provide a measure of uncertainty for the simulated values of wind speed and direction and the variables k and c . We see that assimilating observations of velocity, besides decreasing the error, also reduces the uncertainty of all the variables. Assimilating the observations of concentration substantially reduces the uncertainty of the simulated concentrations, but the uncertainties of the wind speed remain quite large.

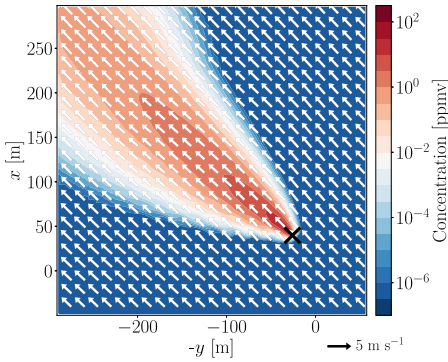
Comparing the results obtained for the two experiments, one can see that assimilating observations of the velocity reduces the error and uncertainty of the flow than the assimilation of observations of concentration. Since dispersion is mainly driven by the flow, the error and uncertainty of concentrations are also significantly reduced through the DA process. Assimilating observations of the velocity allows for a better simulation of the full state of the system and is more efficient than assimilating observations of concentration. The assimilation of both types of observation leads to additionally the reduction in the error of the inflow turbulence.

4.3 Stable Case

Figure 12a, b shows the velocity and concentration fields simulated with the reference boundary conditions in the stable case. Figure 13a shows the departure of the background simulation from this reference, above the containers ($h = 4$ m). The field of velocity error is oriented towards the left, which indicates that the background error is here again mainly along the v -component. In addition, the wind speed is globally overestimated. The erroneous wind speed and direction in the background simulation produce quite different behaviours of the pollutant plume, which can be contrasted to the neutral case: since the wind speed is higher, this leads to an overall underestimation of the concentrations. Moreover, due to the error of wind direction, the pollutant plume far from the source drifts more to the left than the reference plume, which explains why the concentrations are overestimated on the left of the plume (i.e. in line with the buildings) and underestimated on the right.

We can verify in Fig. 13b that the wind speed is overestimated within the container array and that the wind-direction error is globally directed towards the west-south-west (with variations near the source), which means that the background flow is nearly aligned with the streets. As a result, the gas is transported more along the street where the concentrations are thus overestimated left of the plume and underestimated on the right. In the canopy and near the source, the underestimation is more important than overestimation in terms of the amplitude of the errors and extent of the area where concentrations are underestimated.

(a) Reference at $h = 4$ m



(b) Reference at $h = 1$ m

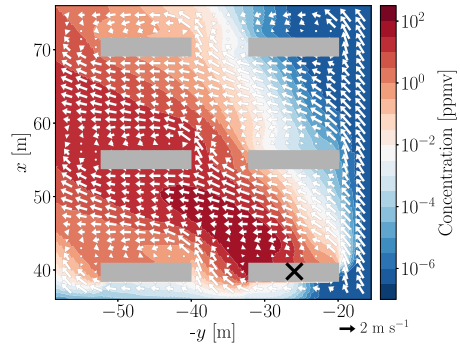
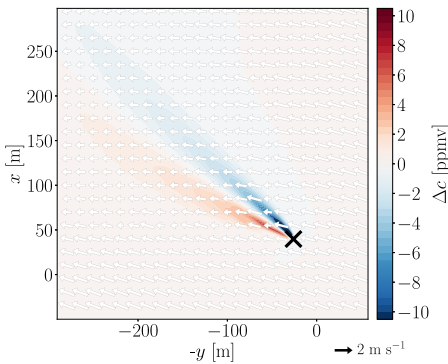


Fig. 12 Reference simulation for the stable case. Horizontal cross-section of the concentration field in a logarithmic scale (colours) and the velocity field (arrows) at (a) $h = 4$ m and (b) $h = 1$ m (magnifying the source)

(a) Background (departure from reference) $h = 4$ m



(b) Background (departure from reference) $h = 1$ m

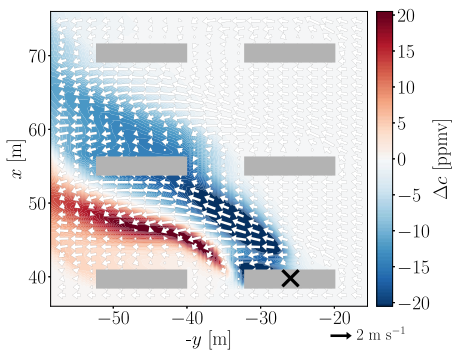


Fig. 13 Departure of the background simulation from the reference in the stable case. The errors of the concentration field (colours) and velocity (arrows) are plotted at (a) $h = 4$ m and (b) $h = 1$ m (magnification near the pollutant source)

We perform the DA experiments starting from the background described above. The optimal BC value is obtained after five IEnKS iterations in experiment W, eight in experiment C, and nine in experiment WC.

Similar to the results obtained in the neutral case, the velocity field of analysis W is closer to the reference than the background and the velocity errors are significantly reduced (see Fig. 14a, b), but small errors of wind speed and direction such that the deviation of the plume to the left persists. The errors for concentrations are reduced by the DA method and are still much smaller than for the background.

Figure 14b shows that, near the source, the same conclusions hold: the velocity field errors are somewhat small such that the pollutant plume is closer to the reference than the background, though not exactly the same. The pattern of concentration errors is similar to that observed for the background, though the amplitude of the errors is reduced and the area where

(a) Analysis W (departure from reference) $h = 4$ m **(b)** Analysis W (departure from reference) $h = 1$ m

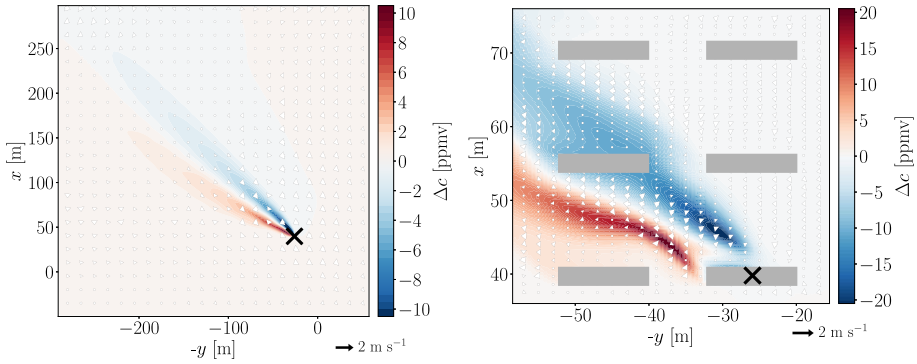


Fig. 14 As Fig. 13 but for the analysis state of experiment W

(a) Analysis C (departure from reference) at $h = 4$ m **(b)** Analysis C (departure from reference) $h = 1$ m

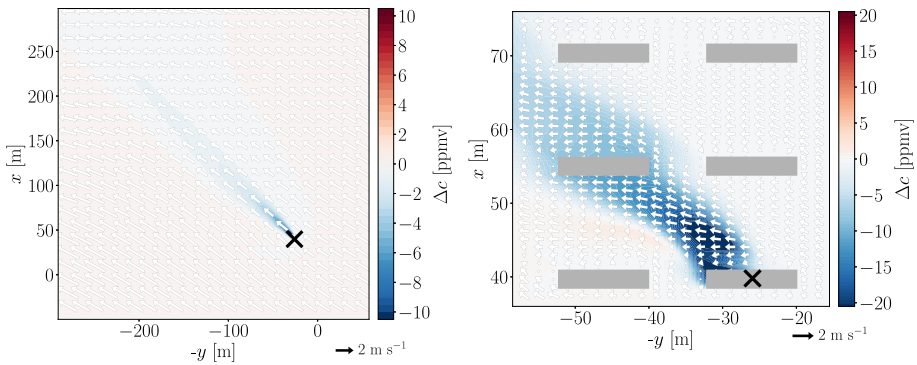


Fig. 15 As Fig. 13 but for the analysis state of experiment C

(a) Analysis WC (departure from reference) $h = 4$ m **(b)** Analysis WC (departure from reference) $h = 1$ m

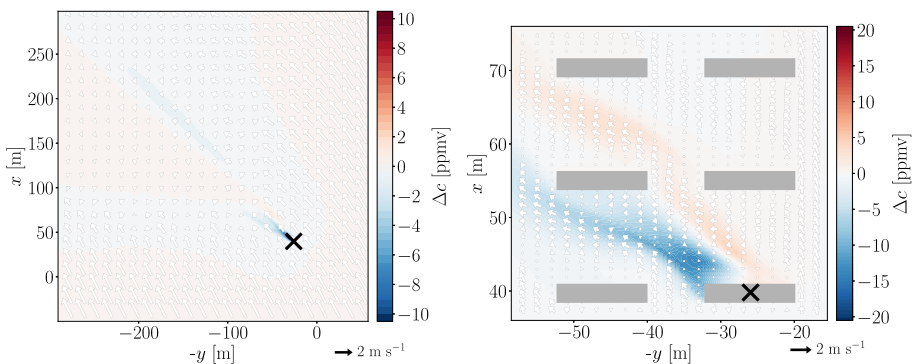


Fig. 16 As Fig. 13 but for the analysis state of experiment WC

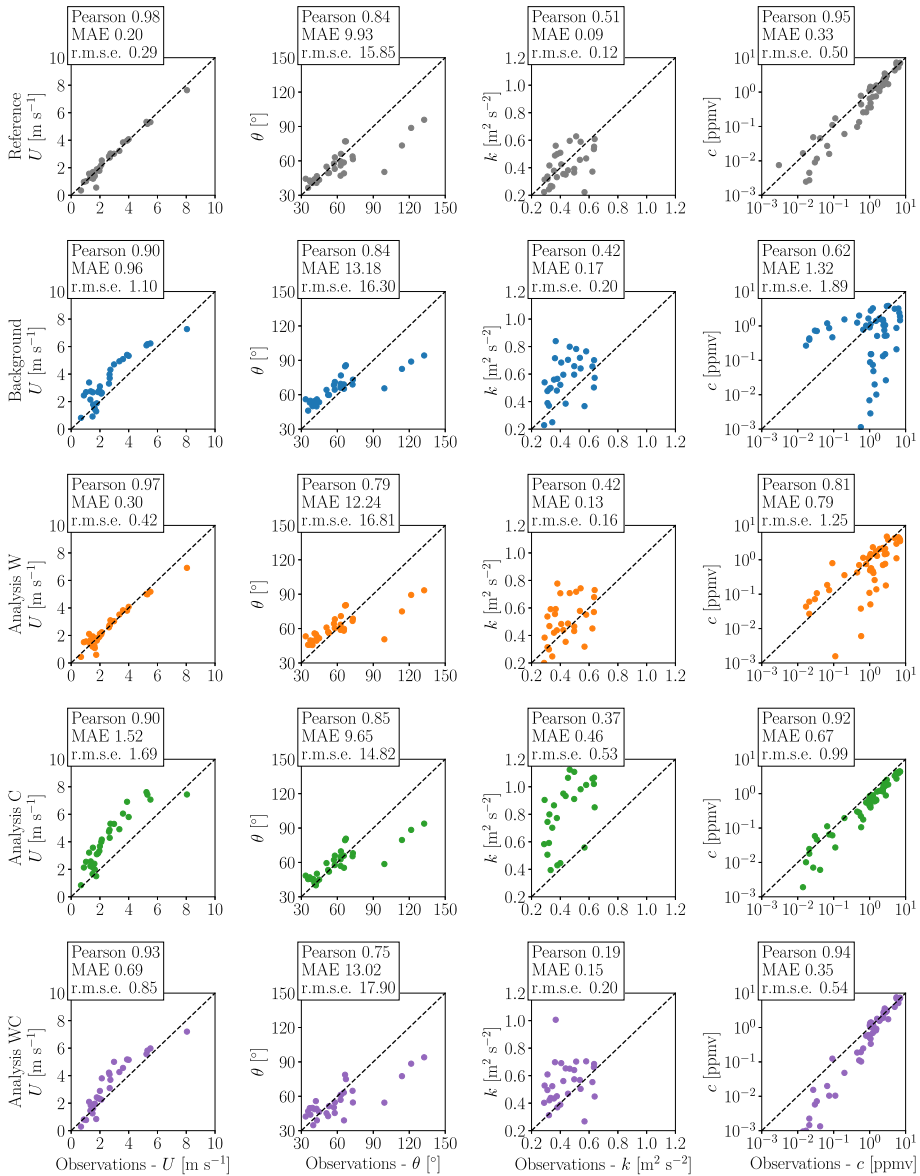


Fig. 17 Simulated versus observed values for the reference simulation (first row), the background (second row), Analysis W (third row), Analysis C (fourth row), and Analysis WC (last row) for the stable case. The scatter plots are shown for wind speed and direction (U , θ), the TKE (k), and the concentration of the tracer gas (c). The Pearson correlation coefficient, the MAE, and the r.m.s. error are also calculated for each variable and simulation. All the available observations are plotted and taken into account in the statistics calculations

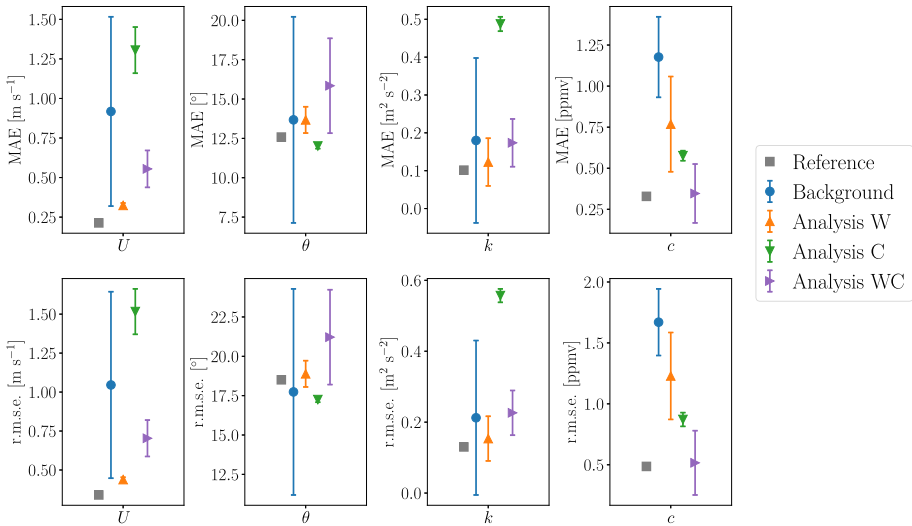


Fig. 18 Mean absolute error and r.m.s. error calculated over the validation observations of U , θ , k , and c within the urban canopy ($h < 2.5$ m) for the reference, background, and analyses W, C, and WC in the stable case. The mean values are calculated as an average over the errors, weighted by the inverse of the number of observations available at the same location. Error bars represent twice the standard deviation of the background and analysis ensembles

concentrations are not well estimated has shrunk. The remaining negative concentration errors in the top-left corner come from the fact that the plume is less spread towards the right in analysis W than in the reference.

Figure 15a shows that the velocity errors for analysis C are aligned with the flow, which means that the wind direction has been corrected. As a result, the pollutant plume is dispersed in the correct direction. However, the wind speed is still overestimated such that the gas is more diluted by the airflow and the concentrations are globally underestimated. The behaviour near the source is also better captured in terms of the location affected by the pollution; however, the concentrations are slightly smaller than the reference. This can be seen in Fig. 15b in the lack of positive concentration errors on the left of the plume, indicating that the direction of plume dispersion is well represented.

Finally, analysis WC shows errors for the velocity field smaller than the background and also smaller than analysis C. Moreover the velocity field errors vanish in the container array (Fig. 16a). Since the velocity is somewhat overestimated, the concentration values are still slightly underestimated after assimilation. We observe in Fig. 16b that, inside the urban canopy, the wind speed is slightly overestimated but the wind direction is in very good agreement with the reference such that the concentration field statistics are very well captured and the concentration errors are very small.

Figures 17 and 18 show a comparison of the results obtained in the stable case for the three experiments with the available observations, with results very similar to those obtained with the neutral case. Note though that the wind direction simulated with background boundary conditions compares quite well with available observations and the mean error is of the same order of magnitude as for reference simulations.

The assimilation of velocity observations in experiment W enables the improved representation of the wind speed and direction. The assimilation of concentration observations

in experiment C efficiently corrects the wind direction but the error for the wind speed is slightly larger than the error in background simulations. In experiment WC, the assimilation of both types of observations leads to similar results as analysis W in terms of wind speed and direction. Note that the errors related to the wind direction for the background simulation are of the same order of magnitude than for reference simulation.

In this case, the TKE is slightly overestimated in the background and the assimilation of velocity observations (experiments W and WC) allows for slight reduction of this error but not the assimilation of concentration observations. As mentioned in Sect. 4.2, the combined effects of dilution and diffusion complicate the assimilation of observations of pollutant concentration alone.

Regarding the concentrations, all analyses show errors that are significantly reduced. Analysis C and WC give better agreement with the observations (see Fig. 17). Figure 17 shows that low values of concentration tend to be underestimated in analysis W, which is consistent with the behaviour of the pollutant plume near the source described above. The best results in terms of concentration are obtained when both types of observations are assimilated (experiment WC).

In the stable case, we can also observe that the DA process leads to a reduction of the uncertainty for all the studied variables (Fig. 18).

4.4 Additional Tests

In addition to the results presented in the two previous sections, we tested the revised version of the IEnKS method with other parametrizations.

In particular, we evaluated the performances of this DA system with different background errors, and performed other DA experiments (not shown) with:

1. the unperturbed background—estimated from the PWIDS, SAMS, and Sound Detection And Ranging (sodar) observations
2. the background perturbed with $\Delta\theta_0 = +10^\circ$,
3. the background perturbed with $\Delta U = 1 \text{ m s}^{-1}$.

In all the cases, assimilating observations of the velocity components helps reduce the error and the uncertainty of the simulated velocity field. Due to the important sensitivity of the plume dispersion to the velocity field, the pollutant plume is generally also better captured. When concentration observations are assimilated, the mean error of concentrations significantly decreases, especially within the urban canopy ($h < 2.5 \text{ m}$). Note that when the perturbation of the wind direction is reversed, as compared with the results shown above ($\Delta\theta = -10^\circ$), the background error of the v -component of velocity is smaller than the error of the u -component and assimilating concentration observations tends to correct the error of u but may worsen the v estimations. Introducing correlations between these two variables, or considering different error statistics for the u and v components in the background-error covariance matrix, could help correct both components of velocity when assimilating concentration observations.

The analyses obtained in the different experiments described above depend on the background, even though this sensitivity is small, which is consistent with the fact that the weight given to the observations is much larger than the one given to the background. This comes from the fact that the eigenvalues of the background error covariance matrix are roughly 25 whereas the eigenvalues of \mathbf{R} are approximately 0.025. As a matter of fact, the part of the

cost function associated with the background

$$\mathcal{J}^b = \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad (11)$$

is practically 10 times smaller than the term associated with the observations

$$\mathcal{J}^o = \frac{1}{2} \|\mathbf{y} - \mathcal{F}(\mathbf{z}^b + \mathbf{A}\mathbf{w})\|_{\mathbf{R}^{-1}}^2. \quad (12)$$

The influence of the radius of vertical correlation (see Eq. 14) has been assessed. A very large radius (typically 50 m) is equivalent to assuming that the error is a bias, nearly constant in the vertical, which is not representative of a typical background error as explained in Appendix 3. As a result, assimilating observations in the canopy helps correct the boundary conditions near the ground and the strong correlation tends to also modify the profile in the upper levels. This may increase the error higher in the atmosphere where the background error is typically small and no additional information is available (observations are more frequent at low altitudes).

5 Conclusions

The IEnKS algorithm has been applied to a case of dispersion modelling in an urban area where buildings affect the flow. These DA experiments also differ from the application of wind-resource assessment presented in Defforge et al. (2018) because the size of the control vector is smaller, though it includes turbulence variables. Moreover, the final outcome of interest here is the concentration of a tracer gas, which is nonlinearly related to the velocity field. The MUST campaign, used here to validate the method, has been widely studied and has the significant advantage of providing numerous observations of velocity and gas concentration.

Among the several trials of gas release performed during the MUST campaign, we selected two trials, corresponding to stable and neutral stability. For each of these two cases, we performed three DA experiments assimilating either velocity or concentration observations, and then both types of observations. For all the experiments, the IEnKS algorithm reduces the error and uncertainty for the assimilated variables: either velocity components or concentration. The departure of simulated concentration fields from validation observations (mean absolute error) is reduced by more than 55% in the neutral case and by nearly 40% in the stable case. Since dispersion is largely driven by the flow, when velocity observations are assimilated, the correction of the flow field through use of the DA method also leads to a better representation of pollutant dispersion. Consequently, the connection between the airflow and dispersion explains that assimilating velocity observations is more efficient for improving the overall dispersion simulation. In contrast, assimilating concentration observations is able to improve the wind direction, which is the parameter that influences the most dispersion. As a result, the concentration field simulated with the analysis boundary conditions is closer to the observations. However, since both the inclusion of the mean wind speed and TKE in the DA algorithm improves the simulation dispersion, assimilating concentration observations only complicates the error reduction for these two variables.

When both types of observations are assimilated, complementary information is added as compared with the case with only velocity observations. As a result, the remaining variables that influence the concentration field, such as turbulence, is also corrected and concentra-

tion errors are even more reduced. Further work would be required to check if assimilating observations of TKE also improves the results.

We assessed the sensitivity of the method to the background error, consistent with the larger confidence placed in the observations compared with the background. While we assume that the observation errors are uncorrelated, this assumption is probably incorrect, especially for the u and v velocity components. Consequently, it would be interesting to perform further sensitivity analyses regarding the observation-error covariance matrix. Moreover, we always worked with ensembles of five members and it would be beneficial to assess the evolution of the performances of the IEnKS algorithm with an increasing ensemble size.

The three experiments presented here correspond to a similar (small) number of observations and one could consider evaluating the sensitivity of the results to the number and the position of the observations. Such an analysis could be particularly beneficial for experimental design of the sensor layout.

Another perspective would be to apply the IEnKS method to the remaining trials performed during the MUST campaign. In particular, the cases 2671934, 2672033 and 2672101 corresponding to very stable atmospheric conditions such that the usual formulae to estimate the vertical profile of boundary conditions cannot be applied. As a result, these cases are known to be particularly difficult to simulate and it would be interesting to try recovering the boundary conditions using the IEnKS algorithm (Kumar et al. 2015).

With the continuous increase in urban monitoring stations, the DA method presented here could be applied for numerous applications. We can expect that, in more complex urban areas, the estimation of background and observation error statistics would be more difficult, while the method might be more sensitive to these parameters. This difficulty apart, the method should still be able to improve the knowledge of boundary conditions and thus the accuracy of pollution maps.

Acknowledgements The CERE laboratory is part of the Institut Pierre-Simon Laplace.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Estimation of the Background-Error Variances and Correlations

Correlation Coefficients

Since the container array is quite small, we assume that the meteorological variables above the urban canopy are homogeneous in the $348\text{ m} \times 348\text{ m}$ region simulated with the *Code_Saturne* model. Consequently, we assume that the velocity observations above the containers within the domain are representative of the values at the border of the domain at the same vertical level.

We thus study the observations provided by:

1. the 6-m towers A, B, C, and D at 6 m,

2. the 16-m towers North and South at 4 m, 8 m, and 16 m,
3. the 32-m tower T at 4 m, 8 m, 16 m, and 32 m.

In order to estimate vertical and horizontal correlations that are representative of the background error, we analyzed the horizontal anomalies of the velocity components (A_u and A_v). The horizontal anomalies are defined as the departure of each observation from the spatial mean computed at the same vertical level:

$$A_{x^j} = x_i^j - \frac{1}{N^j} \sum_{i=1}^{N^j} x_i^j, \quad (13)$$

where x can be u or v , the superscript j refers to a given height above the ground (4 m, 6 m, 8 m, or 16 m), x_i^j is the measurement of the i th instrument at the j th height, and N^j is the number of available observations at this height. The values of standard deviation observed for these horizontal anomalies of velocity and TKE are relatively small, thus confirming the assumption that the meteorological variables are nearly homogeneous over the small domain considered here. It is important to recall that all the trials were selected with south-east to south-west wind directions. Consequently, the asymmetric results obtained for the u - and v -components of velocity can be explained by this bias in the selection of the meteorological conditions for which the releases were performed.

Figure 19 shows the vertical correlations evaluated for the u , v , and k horizontal anomalies from the observations provided by the towers N, S, and T at 4 m, 8 m, and 16 m. The horizontal anomalies of the velocity components u and v show a slight correlation in the vertical.

The correlation coefficient C_{ij} between the two vertical levels i and j of the same profile is assumed to follow the Balgovind law

$$C_{ij} = \left(1 + \frac{d_{ij}}{R}\right) \exp\left(-\frac{d_{ij}}{R}\right), \quad (14)$$

where d_{ij} is the distance between the two levels and R is the correlation length to be determined. As mentioned in the introduction, this law is a simple solution to represent the spatial structure of error statistics (Winiarek 2014). Representing the error statistics well is a significant challenge for the DA method and more sophisticated descriptions could be used, but we chose to use the Balgovind law in all the experiments presented here.

In light of the correlations observed among the horizontal anomalies, and assuming that the climatological correlations underestimate the background error correlations (as explained below), we chose a correlation length $R = 3$ m for the three variables u , v , and k , with the corresponding Balgovind function is shown in Fig. 19. We further assume that the different variables (u , v , and k) are not correlated. Using this hypothesis and the Balgovind correlation function (Eq. 14) with the correlation length given above, we can construct the background-error correlation matrix. This matrix must be multiplied by the standard deviations associated with the i th and j th variables, to get the background-error covariance matrix.

Background-Error Variances

The background-error variances represent the uncertainty of the first estimate. Consequently, these values may depend on the sources of information used to estimate the background boundary conditions. Here, we recreate operational conditions where the first estimate of the boundary conditions is quite poor whereas the observations are very accurate. However, the

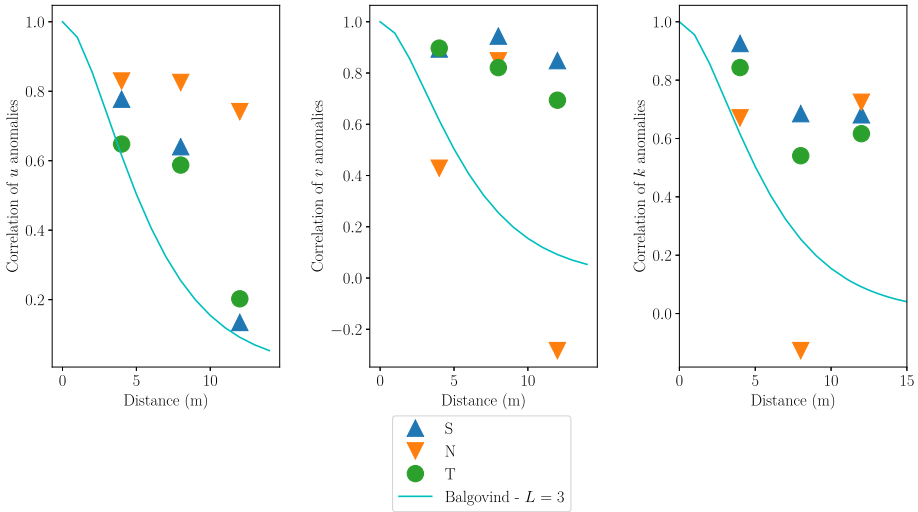


Fig. 19 Vertical correlations estimated from the horizontal anomalies of the parameters u , v , and k for the south (S) and north (N) 16-m masts and the 32-m tower (T). Assuming that the climatological correlations slightly overestimate the background-error correlations, we set the correlation length of the Balgovind function to $R = 3$ m

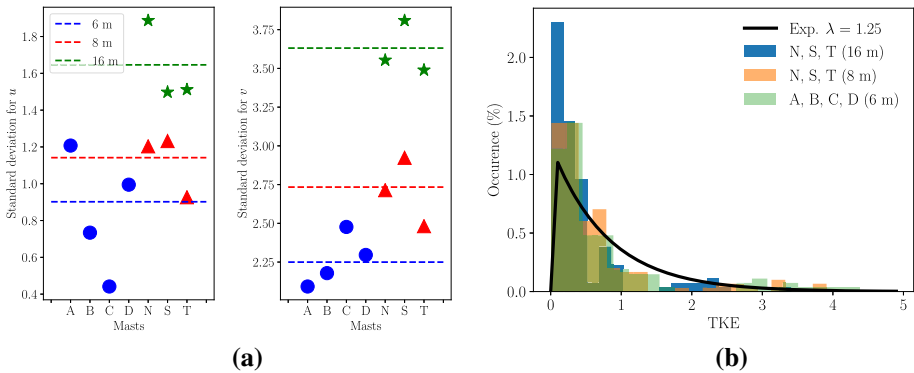


Fig. 20 Statistical analyses of the climatology used to determine the background-error variances. **a** Values of standard deviation estimated from the observations of the velocity components u and v above the canopy, during all the trials. The dashed lines correspond to the standard deviation calculated for each level, merging all the available observations at this height. **b** Probability density function of the observed values of TKE during all the trials compared with the exponential law for the rate parameter $\lambda = 1.25 \text{ s}^2\text{m}^{-2}$

MUST case is very specific and does not correspond to operational conditions. Indeed, the trials were performed in particular meteorological conditions, not necessarily representative of the full climatology in this region. Moreover, the container array is installed in the middle of a desert area, such that the meteorological conditions outside the domain are not perturbed by any geometrical features and remain spatially homogeneous. As a result, we assume that the climatological variations, observed over all the trials periods, for the velocity components and TKE slightly underestimate the background-error variances. Using the climatological analyses over all trials (see Fig. 20a) and expert judgement, we set the background-error variance for u , v , and k to $\sigma_{uv}^b = \sigma_{\gamma}^b = 5$.

Appendix 2: Coefficients of the Observation-Error Covariance Matrix

The diagonal terms of the observation error covariance matrix are of the form

$$\mathbf{R}_i = (\sigma^o)^2 N_{xy}. \tag{15}$$

Since the extra-diagonal terms of \mathbf{R} are assumed zero (i.e., the observations are not correlated) with this form for \mathbf{R} , (Eq. 10), the part of the cost function associated with the observations reads

$$\mathcal{J}^o = \frac{1}{2} \|\mathbf{y} - \mathcal{F}(\mathbf{z})\|_{\mathbf{R}^{-1}}^2 \tag{16a}$$

$$= \sum_{i=1}^p (\mathbf{R}^{-1})_{ii} (\mathbf{d}\mathbf{y}_i)^2 \tag{16b}$$

$$= \sum_j \frac{\sum_{i=1}^{N_{x_j, y_j}} (\mathbf{d}\mathbf{y}_i)^2}{(\sigma^o)^2 N_{x_j, y_j}}, \tag{16c}$$

where $\mathbf{d}\mathbf{y}$ is the innovation term and the sum over j represents the different horizontal positions where observations are available. Consequently, the cost function associated with the observations is a sum of innovations, averaged for each horizontal position over the different measurements there.

Appendix 3: Background Boundary Conditions

Estimation from PWIDS, SAMS, and Sodar Observations

The first estimate of the profile of boundary conditions for the wind speed U is assumed to follow the semi-logarithmic profile given by the Monin–Obukhov similarity theory (Stull 1988)

$$U(z) = \frac{U_L}{\kappa} \left(\log \left(\frac{z + z_0}{z_0} \right) + 5 \frac{z}{L} \right), \tag{17}$$

where $\kappa = 0.4$ is the von Kármán constant, z_0 is the roughness length which is set to 0.04 m in the MUST domain (Yee and Biltoft 2004), and L is the Obukhov length. The local friction velocity U_L is estimated from the boundary-layer height h_{ABL} and the surface friction velocity u_*

$$U_L = u_* \left(1 - \frac{z}{h_{\text{ABL}}} \right). \tag{18}$$

In Garratt (1994), the height of the boundary layer is estimated in neutral conditions as

$$h_{\text{ABL}} = 0.2u_*/|f|, \tag{19}$$

and in stable conditions as

$$h_{\text{ABL}} = 0.4\sqrt{u_*L/|f|}, \tag{20}$$

where f is the Coriolis frequency, equal to $f = 9.4 \times 10^{-5} \text{ rad s}^{-1}$ here.



Fig. 21 Map of the site around the MUST array. The light orange pins represent the corners of the container array. The location of the SAMS #8 (red), the PWIDS M6 (blue), and the sodar (green) are also shown

We also assume that the profile of TKE is consistent with the Monin–Obukhov similarity theory

$$k = \frac{U_L^2}{\sqrt{C_\mu}}, \quad (21)$$

$$\epsilon = \frac{U_L^3}{\kappa} \left(\frac{1}{z} + \frac{4}{L} \right). \quad (22)$$

This scheme has been proved to be well suited for the MUST case (Milliez 2006).

In order to determine the friction velocity u_* , we consider the background observations provided by the instruments shown in Fig. 21 and detailed in Table 1 for the two cases. We then fit a semi-logarithmic velocity profile to the observations to determine the value of u_* . The wind direction is assumed constant over the vertical profile, and is obtained as an average over the available observations. For the neutral case, we find $u_* = 0.57 \text{ m s}^{-1}$, which corresponds to an atmospheric boundary-layer height of $h_{ABL} = 1.21 \text{ km}$, and the mean wind direction is equal to -42.2° (with a standard deviation among the observations of 2.6°). For the stable case, the estimated surface stress $u_* = 0.40 \text{ m s}^{-1}$, $h_{ABL} = 298 \text{ m}$, the wind direction is estimated as 45.4° (with a standard deviation among the observations of 4°). Figure 22a, b show the profiles of wind speed and direction estimated from the observations (red dots) for the neutral case.

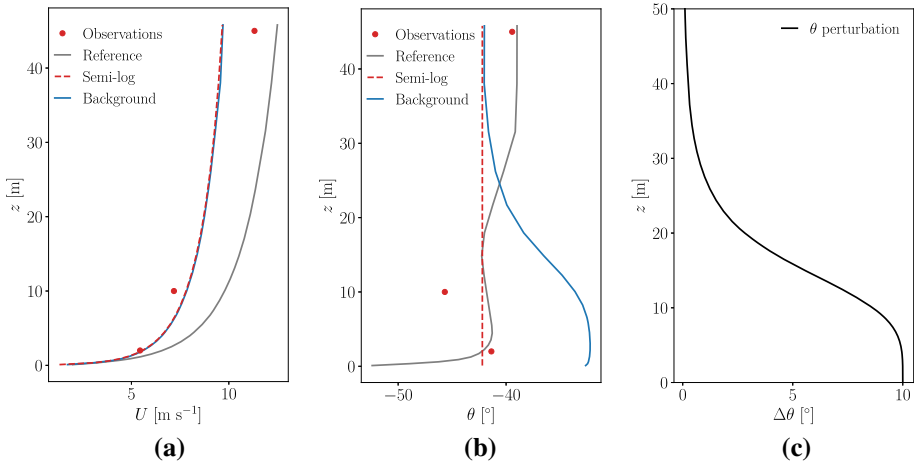


Fig. 22 Construction of the background boundary conditions: Vertical profiles of **a** wind speed (U) and **b** wind direction (θ) for the neutral case, showing the reference boundary conditions, the semi-logarithmic profile reconstructed from the PWIDS, SAMS, and sodar observations, and the background boundary conditions (obtained by perturbing the semi-logarithmic profile) are shown. **c** Perturbation function of wind direction, decreasing with altitude, which is added to the constant profile of θ obtained from the average over the observed values

Perturbation of the Boundary Conditions

As previously mentioned, the MUST case is very specific since the area surrounding the container array is a flat desert and, in the absence of obstacles, the flow field is spatially homogeneous. As a matter of fact, the vertical profiles of velocity and TKE observed by the PWIDS, SAMS, and sodar stations—even though these instruments are quite far from the container array—is a very good estimate of the boundary conditions. In fact, when the prescribed boundary conditions are estimated as explained above, the mean velocity-field error inside the domain—computed as the MAE, compared to the available flow observations—is smaller than 1.1 m s^{-1} . This error is smaller than what is usually encountered in operational studies.

In order to make this study representative of an operational application, we empirically perturb the boundary conditions according to the function

$$\delta X = \delta X_0 \frac{\alpha}{(h/h_L)^4 + \alpha}, \tag{23}$$

where δX is the perturbation for the variable X , δX_0 is the maximum perturbation near the ground, h is the elevation, h_L is the level at which the perturbation is equal to $p_L \delta X_0$, and $\alpha = p_L / (1 - p_L)$ (Fig. 22c). This function perturbs the profile nearer to the ground than in the upper layers, which is representative of a typical background error. In operational studies, the background is generally estimated from mesoscale simulations or from observations outside the domain under scrutiny.

Mesoscale models generally provide poorer estimations of the meteorological fields near the surface, where the local effect of topography, which is generally inaccurately represented in such model, is greater. As a result, if mesoscale results are used to estimate the background, it is likely that the background error is larger near the ground than in the upper layers. Similarly, the variability in time and space of the meteorological variables is greater in the lower

levels than in the uppermost ones. Consequently, the observations provided by instruments a few hundred metres away from the domain of interest can generally be considered as a good estimation of the boundary conditions above a few tens of metres from the ground (i.e. largely above the canopy, even for cases where geometrical features perturb the flow field outside the studied domain). However, in the lower levels, the use of perturbed canopy measurements as boundary conditions is very likely to be wrong.

In order to be consistent with typical background errors, the perturbation that we apply is larger near the ground and vanishes to zero with altitude, see (Eq. 23). Here, we perturb only the wind direction with different amplitudes ($\Delta\theta_0 = \pm 10$) and with $p_L = 0.01$ and $h_L = 50$ m (see Fig. 22c).

References

- Albriet B, Sartelet K, Lacour S, Carissimo B, Seigneur C (2010) Modelling aerosol number distributions from a vehicle exhaust with an aerosol CFD model. *Atmos Environ* 44(8):1126–1137
- Archambeau F, Méchitoua N, Sakiz M (2004) Code Saturne: a finite volume code for the computation of turbulent incompressible flows-industrial applications. *Int J Finite Volumes* 1:1–62
- Asch M, Bocquet M, Nodet M (2016) Data assimilation: methods, algorithms, and applications. Society for Industrial and Applied Mathematics, Philadelphia
- Auroux D, Blum J (2008) A nudging-based data assimilation method: the Back and Forth Nudging (BFN) algorithm. *Nonlinear Process Geophys* 15:305–319
- Bahlali M (2018) Adaptation de la modélisation hybride eulérienne/lagrangienne stochastique de Code_Saturne à la dispersion atmosphérique de polluants à l'échelle micro-météorologique et comparaison à la méthode eulérienne. Ph.D. thesis, Université Paris-Est. <http://www.theses.fr/2018PESC1047/document>
- Bahlali ML, Dupont E, Carissimo B (2019) Atmospheric dispersion using a lagrangian stochastic approach: application to an idealized urban area under neutral and stable meteorological conditions. *J Wind Eng Ind Aerodyn* 193(103):976
- Balgovind R, Dalcher A, Ghil M, Kalnay E, Balgovind R, Dalcher A, Ghil M, Kalnay E (1983) A stochastic-dynamic model for the spatial structure of forecast error statistics. *Mon Weather Rev* 111(4):701–722
- Benamrane Y, Wybo JL, Armand P (2013) Chernobyl and Fukushima nuclear accidents: what has changed in the use of atmospheric dispersion modeling? *J Environ Radioact* 126:239–252
- Bertino L, Evensen G, Wackernagel H (2003) Sequential data assimilation techniques in oceanography. *Int Stat Rev* 71(2):223–241
- Biltoft CA (2001) Customer report for mock urban setting test. DPG Document Number 8-CO-160-000-052 Prepared for the Defence Threat Reduction Agency
- Björck Å (1996) Numerical methods for least squares problems. Society for Industrial and Applied Mathematics, Philadelphia
- Blocken B, Tominaga Y, Stathopoulos T (2013) CFD simulation of micro-scale pollutant dispersion in the built environment. *Build Environ* 64:225–230
- Bocquet M, Sakov P (2014) An iterative ensemble Kalman smoother. *Q J R Meteorol Soc* 140:1521–1535
- Cohn SE (1997) An introduction to estimation theory (gtspecial issue\data assimilation in meteorology and oceanography: theory and practice). *J Meteorol Soc Jpn Ser II* 75(1B):257–288
- Davoine X, Bocquet M (2007) Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport. *Atmos Chem Phys* 7(6):1549–1564
- Defforge CL, Carissimo B, Bocquet M, Armand P, Bresson R (2018) Data assimilation at local scale to improve CFD simulations of atmospheric dispersion: application to 1D shallow-water equations and method comparisons. *Int J Environ Pollut* 64(1/3):90–109
- Defforge CL, Carissimo B, Bocquet M, Bresson R, Armand P (2019) Improving CFD atmospheric simulations at local scale for wind resource assessment using the iterative ensemble Kalman smoother. *J Wind Eng Ind Aerodyn* 189:243–257
- Garratt J (1994) Review: the atmospheric boundary layer. *Earth Sci Rev* 37(1–2):89–134
- Hanna SR, Brown MJ, Camelli FE, Chan ST, Coirier WJ, Hansen OR, Huber AH, Kim S, Reynolds RM, Hanna SR, Brown MJ, Camelli FE, Chan ST, Coirier WJ, Hansen OR, Huber AH, Kim S, Reynolds RM (2006) Detailed simulations of atmospheric flow and dispersion in downtown Manhattan: an application of five computational fluid dynamics models. *Bull Am Meteorol Soc* 87(12):1713–1726

- Holmes N, Morawska L (2006) A review of dispersion modelling and its application to the dispersion of particles: an overview of different dispersion models available. *Atmos Environ* 40(30):5902–5928
- Kalnay E (2003) Atmospheric modeling, data assimilation, and predictability. Cambridge University Press, Cambridge
- Kovalets IV, Tsiouri V, Andronopoulos S, Bartzis JG (2009) Improvement of source and wind field input of atmospheric dispersion model by assimilation of concentration measurements: Method and applications in idealized settings. *Appl Math Model* 33(8):3511–3521
- Krysta M, Bocquet M, Sportisse B, Isnard O (2006) Data assimilation for short-range dispersion of radionuclides: an application to wind tunnel data. *Atmos Environ* 40(38):7267–7279
- Kumar P, Ketzel M, Vardoulakis S, Pirjola L, Britter R (2011) Dynamics and dispersion modelling of nanoparticles from road traffic in the urban atmospheric environment—a review. *J Aerosol Sci* 42(9):580–603
- Kumar P, Feiz AA, Singh SK, Ngae P, Turbelin G (2015) Reconstruction of an atmospheric tracer source in an urban-like environment. *J Geophys Res Atmos* 120(24):12589–12604
- Liu Y, Haussaire JM, Bocquet M, Roustan Y, Saunier O, Mathieu A (2017) Uncertainty quantification of pollutant source retrieval: comparison of Bayesian methods with application to the Chernobyl and Fukushima Daiichi accidental releases of radionuclides. *Q J R Meteorol Soc* 143(708):2886–2901
- Milliez M (2006) Modélisation micro-météorologique en milieu urbain: dispersion des polluants et prise en compte des effets radiatifs. Ph.D. thesis, ENPC. <https://pastel.archives-ouvertes.fr/pastel-00004042/document>
- Milliez M, Carissimo B (2007) Numerical simulations of pollutant dispersion in an idealized urban area, for different meteorological conditions. *Boundary-Layer Meteorol* 122(2):321–342
- Milliez M, Carissimo B (2008) Computational fluid dynamical modelling of concentration fluctuations in an idealized urban area. *Boundary-Layer Meteorol* 127(2):241–259
- Mons V, Margheri L, Chassaing JC, Sagaut P (2017) Data assimilation-based reconstruction of urban pollutant release characteristics. *J Wind Eng Ind Aerodyn* 169:232–250
- Robins A (2003) Wind tunnel dispersion modelling some recent and not so recent achievements. *J Wind Eng Ind Aerodyn* 91(12–15):1777–1790
- Sakov P, Oliver DS, Bertino L (2012) An iterative EnKF for strongly nonlinear systems. *Mon Weather Rev* 140:1988–2004
- Sousa J, García-Sánchez C, Górlé C (2018) Improving urban flow predictions through data assimilation. *Build Environ* 132(February):282–290
- Srebric J, Vukovic V, He G (2008) CFD boundary conditions for contaminant dispersion, heat transfer and airflow simulations around human occupants in indoor environments. *Build Environ* 43(3):294–303
- Stull RB (1988) An introduction to boundary layer meteorology. Springer, Dordrecht
- Winiarek V (2014) Dispersion atmosphérique et modélisation inverse pour la reconstruction de sources accidentelles de polluants. Ph.D. thesis, Université Paris-Est. <https://pastel.archives-ouvertes.fr/tel-01004505/document>
- Winiarek V, Bocquet M, Saunier O, Mathieu A (2012) Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant. *J Geophys Res Atmos* 117(D5):D05122
- Yee E, Biltoft CA (2004) Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles. *Boundary-Layer Meteorol* 111(3):363–415