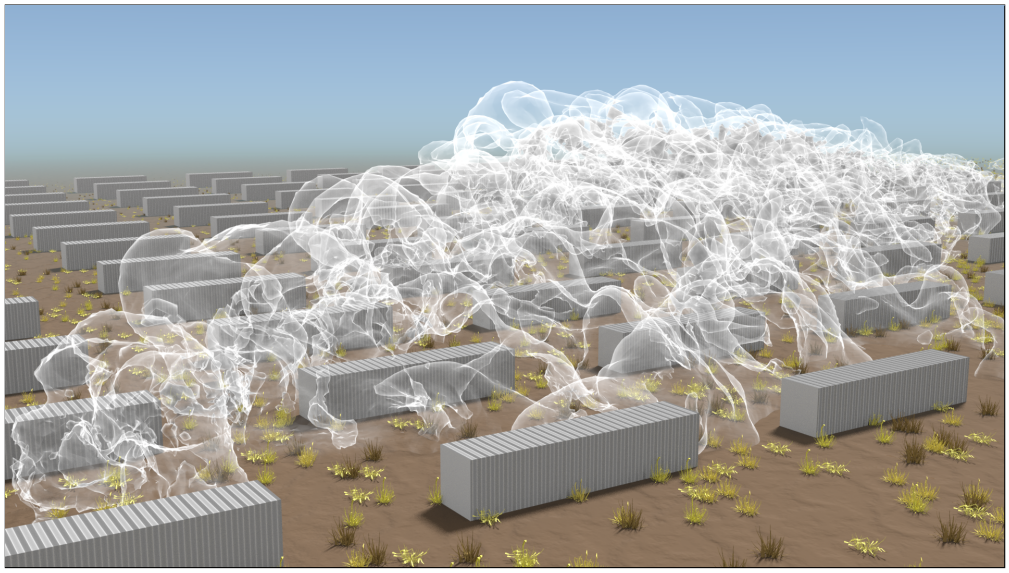
Large-eddy simulations of the MUST trial 2681829 using AVBP, Meso-NH and YALES2

Laetitia Thouron, Mélanie Rochoux, Thomas Jaravel, Olivier Vermorel

Large-eddy simulations of the MUST trial 2681829 using AVBP, Meso-NH and YALES2

Laëtitia Thouron, Mélanie C. Rochoux, Thomas Jaravel and Olivier Vermorel

Technical report, TR-CFD-20-14, CERFACS, Toulouse, France



Blender visualisation of the large-eddy simulation/LES of the MUST trial 2681829 using AVBP

This document is a technical note to complement the article manuscript by Rochoux et al. [33]. Our objective is to provide an extensive comparison of Large-Eddy Simulations (LES) for micro-scale dispersion of air pollutants in the framework of the Mock Urban Setting Test (MUST) experiment. This is done through a validation against experimental data and through a detailed model-to-model comparison. For this purpose, we investigate the capability of LES to capture the unsteady short-to-medium-range plume dynamics and dispersion in near-neutral boundary layer conditions using three different LES codes: AVBP, Meso-NH (including an immersed boundary method/IBM) and YALES2. We also analyse the sensitivity of the LES results to different physical and numerical choices (e.g. model equations, computational grids, numerical schemes, physical assumptions).

This work is part of current CERFACS' efforts to design safety CFD (Computational Fluid Dynamics) capabilities with applications to environment (e.g. air pollution dispersion, wildland fires).

INTRODUCTION

The general objective of this work is to assess the **robustness of Large-Eddy Simulation (LES)** for simulating near-field air pollution problem at very high resolution (i.e. at submeter scale) in the atmospheric boundary layer and provide an estimation of the LES multi-model variability for a well-known test case (Mock Urban Setting Test/MUST). The reader shall refer to Rochoux et al. [33] for the scientific presentation and analysis of the LES results; the objective here is to further **justify scientific and technical choices** made in the study.

The present document reports several numerical tests that were carried out to analyse the sensitivity of the results to numerical choices done at both the model integration and postprocessing stages with the following LES codes: **AVBP** [14, 36], **MesoNH** [18, 19] (including an Immersed Boundary Method/IBM [1, 2] and referred to as MNH-IBM in the following) and **YALES2** [23, 27]. These three codes are references in their communities, CFD and combustion for AVBP and YALES2 on the one hand, atmosphere for MNH-IBM on the other hand. Over the years, they have been applied to a wide range of applications and validated against multiple observational datasets. Still, none of the three codes has been used here in its classic application framework: the MUST test case is the first near-real case simulated using the IBM in MNH-IBM; AVBP and YALES2 are not usually applied to environmental fluid flows in open areas.

Picture of the MUST experiment and shipping containers in Utah's West Desert [3]

Our primary objective is **model validation** [4, 9, 13, 25], i.e. to investigate whether the LES solvers routinely used at CERFACS are able to simulate in an accurate way, the plume dynamics following a gas release from a point source and its micro-scale dispersion through an idealized urban canopy.

Our scientific objective is **exhaustive model-to-model comparison**, i.e. to use the pool or ensemble of LES models (the ensemble size is small, on the order of 10) to estimate multi-model variability [10, 30], which could be related in the future to the natural variability associated with large-scale meteorological forcing, in order to quantify the total uncertainty in such LES model predictions.

Structure of the document:

- 1) The MUST case study is briefly presented.
- 2) The pool of LES model configurations is presented, with a focus on AVBP specificities to justify model configurations (e.g. artificial compressibility, boundary conditions, mesh properties) that are not presented in Rochoux et al. [33].
- 3) LES results are reported with a more exhaustive list of plots and statistical metrics than in Rochoux et al. [33] for completeness of the model-to-model comparison.
- 4) Explanations on the structural model uncertainty estimation approach are given.
- 5) A preliminary step towards the aleatory uncertainty estimation is reported to pave the way towards quantifying the impact of the large-scale meteorological forcing on micro-scale meteorology in future work.

Lexicon:

- CFD/Computational Fluid Dynamics
- CFL/Courant-Friedrichs-Lewy
- dPID/Digital Photo-Ionization Detector
- FAC2/Fraction of predictions within a fACtor of TWO of observations
- FB/Fractional Bias
- IBM/Immersed Boundary Method
- LES/Large-Eddy Simulation
- LW/Lax-Wendroff
- MG/Geometric Mean bias
- MUST/Mock Urban Setting Test
- MME/Multi-Model Ensemble
- NMSE/Normalized Mean Square Error
- PPM/Piecewise Parabolic Method
- PGS/Pressure Gradient Scaling
- TKE/Turbulent Kinetic Energy
- TTG/Two-step Taylor-Galerkin
- VG/Geometric Variance
- WALE/Wall-Adaptative Local Eddy-Viscosity
- WENO/Weight Essential Non Oscillatory

1) THE MOCK URBAN SETTING TEST

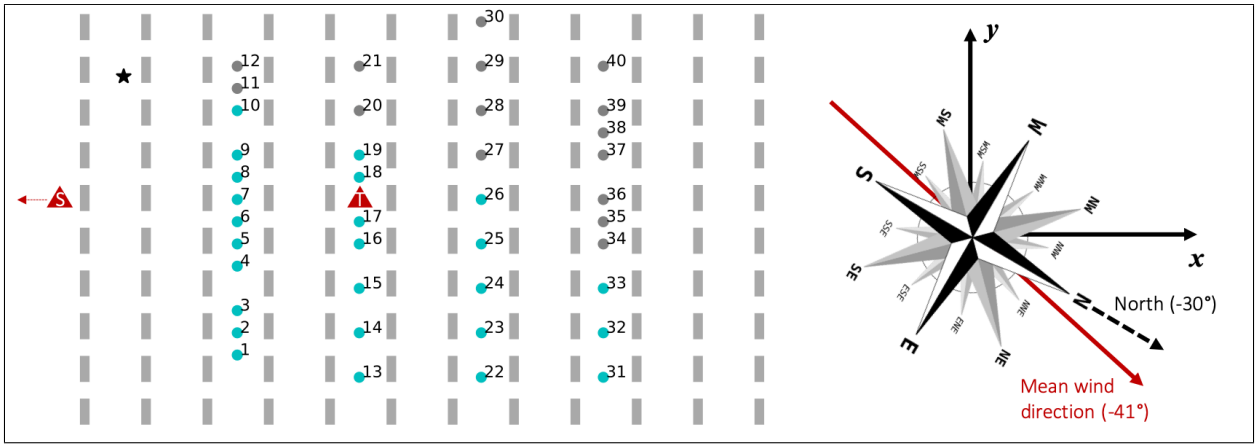


Figure 1: *Map of the shipping containers (grey rectangles) and the 40 tracer concentration probes located at 1.6-m high (circle symbols) in the MUST case study. Grey circles correspond to probes where the observed tracer concentration during the MUST trial 2681829 is below the instrument detection limit (0.04 ppm): these probes were removed from the dataset to compute the performance metrics; only 25 probes (blue circles) were used. Red triangles correspond to instrumented towers, the upstream tower S and the central tower T. A windrose is shown to indicate the orientation of the cardinal directions with respect to the array x- and y-axes: the mean wind direction of the MUST trial 2681829 is indicated in red.*

1.1) EXPERIMENTAL TEST SITE AND POINT-SOURCE EMISSION

Our test case is the MUST field campaign [3, 38] corresponding to a gas tracer emitted as a point source in an idealized urban environment in a continental meteorology in late September 2001 at Horizontal Grid on the U.S. Army Dugway Proving Ground, located in the Great Basin Desert of north-western Utah (USA).

The test site is mainly flat (slope of approximately 0.0005 towards the south). It is homogeneously covered with a mixture of sparse greasewood and sagebrush[1] ranging in height from about 0.4 to 0.75 m. The average roughness length (z_0) was estimated to 0.045 m ($\pm$ 0.005 m). The north direction (dashed arrow) corresponds to the angle -30° in the x-y frame presented in in Figure 1. This orientation was chosen *to take advantage of the prevailing wind directions at the test site in the sector from south-south-west to south-east, resulting in a flat and homogeneous upwind fetch of more than 10 km* according to Yee and Biltoft [38].

The urban environment is made of an array of containers: conex containers were regularly arranged in a 12-by-10 array in the x-y frame shown in Figure 1. Each container is 12.2-m long, 2.42-m wide, and 2.54-m high (H). The mean distances between two rows and two lines (corresponding to the horizontal *x*-direction and vertical *y*-direction in Figure 1) are on average 12.9 m (W) and 7.9 m, respectively. This geometrical arrangement can be classified as an avenue canyon or as a semi-open area with aspect ratio equal to 0.2 (i.e. the height-to-width ratio H/W). The overall size of the container array is 93 m by 171 m. Note that some small geometrical irregularities were present in the way the containers were arranged during the experiment; we still consider a regular case as for instance in Ref. [26].

The emitted gas is propylene ($\mathrm{C_3H_6}$), which is advantageous for included cost, availability, safety and health, and compatibility with the instruments. It is also representative of gases used in chemical industry. Propylene can be considered as non reactive and is therefore represented as a tracer. It was released through the container array during 15 minutes, at different locations and for different atmospheric conditions (several wind directions, wind speeds and atmosphere states) across the multiple (63) MUST trials.

Measurements of mean velocity and turbulence were obtained within and above the container array. Vertical profiles were obtained from the 32-m tower T located near the centre of the container array (red triangle inside the container array in Figure 1). This central tower was instrumented with four 3-D sonic anemometers located at 4, 8, 16 and 32 m above the ground level. To track the tracer within the container array, a number of measuring stations (40 dPIDs/digital photo-ionization detectors represented as circles in Figure 1) were

[1] Typical shrubland found in California and the northern part of Mexico shaped by a Mediterranean climate and wildfire, featuring summer-drought-tolerant plants (greasewood stands for “chaparral" and sagebrush for “armoise" in French)

placed in-between the containers at 1.6-m high, they were arranged along four sampling lines (aligned with the vertical *y*-direction in Figure 1). These stations referred to as "probes" hereafter recorded the time series of tracer concentration (in ppm). Note that we only have access to these 40 horizontal probes in the present analysis. Vertical profiles of concentration statistics at the central tower T and the 6-m masts A-B-C-D were not used. Mean velocity and turbulence measurements at the masts A-B-C-D were also not used. Future work could include these data in the analysis.

1.2) SELECTED EXPERIMENTAL TRIAL

63 MUST trials were performed experimentally. In our work, we focus on one trial referred to as 2681829, which is part of the 21 trials retained by Yee and Biltoft (2004) [38] for a detailed analysis. We made the choice of studying a single trial due to the high overall computational cost of the study and the complexity of the LES multi-model result analysis.

The characteristics of the selected trial are summarized in Table 1. This case is a configuration with near-neutral atmospheric conditions (i.e. afternoon transition from unstable to stable conditions), characterized by a high value of the Obukhov length L_o (> 2,500 m at the 4-m level of the central tower T), no latent and sensible heat fluxes, and a weak influence of buoyancy. This case with ineffective thermal turbulence is thus particularly relevant to study mechanical turbulence such as the container impact on the incident flow and in particular the plume deflection from the inlet wind direction within the container array.

Intermittent wind conditions were recorded at the upstream tower S (red triangle outside of the container array in Figure 1) over the 15 minutes of the trial. Figure 2 presents the time-evolution of the near surface-wind conditions during the first 200-s time period of interest: the wind speed and the wind direction range from -83 to -12° and from 3.2 to 13.9 m s^1 at 4-m high at tower S. The (time-averaged) mean wind speed (u_4) and wind direction (α_4) at 4 m high are respectively 7.93 m s^{-1} and -41°. These conditions can be considered as moderate (neither too calm nor too strong), with a wind mainly blowing from the south-southwest direction. Note that these mean wind conditions were recorded at different heights (1.15, 4, 8 and 16 m) of the upstream tower S; they are used to derive the initial and inlet wind fields for the LES models (Figure 4).

Note that in Ref. [38], a 200-s time period was extracted from the 15-minute continuous release during the trial corresponding to the least variation in the mean wind speed and direction recorded from the upwind mast anemometers. This was done in order to remove unstationarity from the data. This "optimal" time period is between 300 and 500 s according to unpublished information (private communication with Bertrand Carissimo/CEREA and Tim Nagel/CNRM). Future work could include comparing the simulation data against the choice of the optimal time sequence.

The tracer was released and continuously sampled at a source point (at 1.8 m high) using a mass flow controller to maintain a constant flow rate (225 L min^{-1} with a 2 % uncertainty) over a period of 15 minutes. The source location is indicated by a black star symbol in Figure 1. According to Yee and Biltoft [38], the release can be considered as passive at this flow rate level.

Table 1: *Main characteristics of the MUST trial 2681829 [38]: α_4 and u_4 are the mean wind direction and the mean horizontal wind speed at the 4-m level of the upstream tower S; L_o is the Obukhov length at the 4-m level of the central tower T; Q is the tracer release rate; and z_s is the source height.*

Trial	Start local time UTC	α_4 (°)	u_4 (m s^{-1})	L_o (m)	Q (L min^{-1})	z_s (m)
2681829	2001/09/25 1830	-41	7.93	28,000	225	1.8

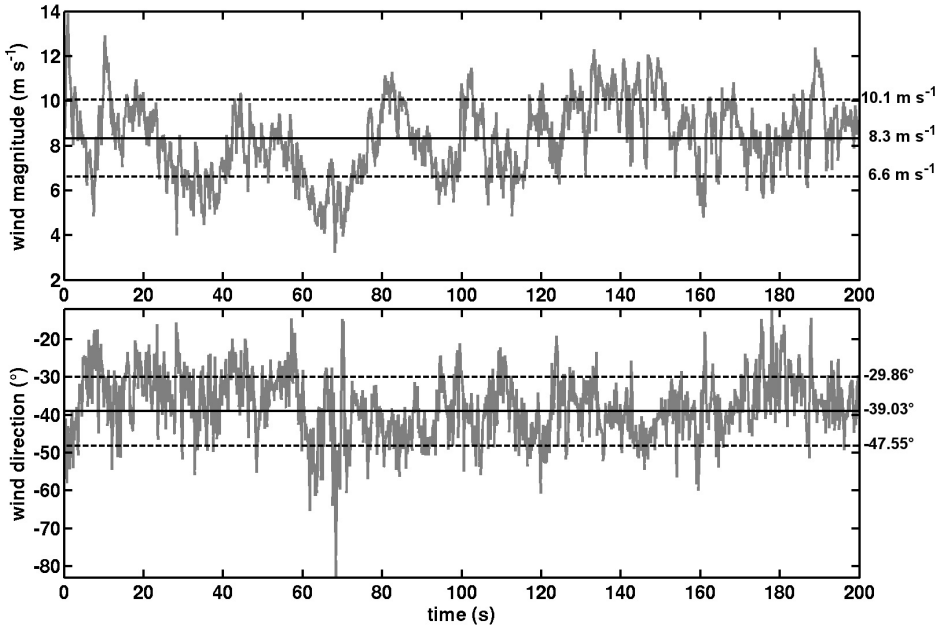


Figure 2: *Time series of wind speed ($m\,s^{-1}$, top panel) and wind direction ($^{\circ}$, bottom panel), at the 4-m level of the upstream tower S, over the first 200 s of the MUST trial 2681829 (grey solid lines); the wind direction is defined with respect to the x-axis of the container array (-30° corresponds to a wind blowing from the south to the north direction). The mean values are indicated in black solid lines; the mean values more or less the standard deviation are indicated in black dashed lines.*

1.3) OBSERVATIONAL DATASET POSTPROCESSING

a) Data acquisition

Probes provide "raw" data, which are to be processed to be compared to the LES model outputs. One difference between the observed and simulated datasets relates to the acquisition time (i.e. the time period at which data are saved); the acquisition frequency of the observations is ten times higher than that of the model outputs. This difference is to remember when considering nominal values. However, the evaluation statistical metrics proposed by Chang and Hanna [6] and used in the present analysis consider (time-averaged) mean quantities so that the difference in acquisition time is not an issue. A second difference relates to the initial time of data acquisition. For a proper model evaluation, it is of first importance to synchronise the initial time of observed data with that of the simulated times (when a signal starts to be observed at the simulated probes). Instruments have an initialization phase of about 61 seconds. LES have an initialization stage of about 30 seconds: 15 seconds are needed to spin-up the flow field and establish the container-induced turbulence, and 15 additional seconds are then needed to establish the plume in the computational domain. Table 2 gives the actual time offsets to be subtracted from the observations and for each LES code. Note that a 15-s spin-up is indicated for YALES2, since YALES2 simulations were initialized at this preliminary stage using an AVBP solution to gain from computational time (there is a tool to map an AVBP solution onto a YALES solution). Note also that the same synchronisation is applied for all probes of a given simulated configuration.

Table 2: *Time offsets between observations and LES models*

Offsets	Probes	MNH-IBM	YALES2	AVBP
Sensor initialization phaze (s)	61	–	–	–
Wind spin-up (s)	–	15	–	15
Tracer spin-up (s)	15	15	15	15
Total waiting time (s)	76	30	15	30

b) Instrument limit detection

The instrument limit detection for each probe ranges between 0.04 and 1,000 ppm according to Biltoft [3]. As a consequence, every probe that never records above 0.04 ppm at any instantaneous time of the 200-s time sequence is removed from the present analysis. Following this principle, 15 probes record measurements that are below the instrument limit detection. They are indicated using grey markers in Figure 1 and correspond to probes numbered 11, 12, 20, 21, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39 and 40. Thus, the total number of probes considered in the present analysis is 25 (instead of 40).

Milliez and Carissimo [26] reported that the CFD model performance could be improved by considering a minimum threshold of 0.1 ppm. To address this point, we compare in the present document the results obtained for a threshold of 0.04 ppm and 0.1 ppm (Table 12).

Note that a filter is applied for time steps with values below the lower instrument limit detection: the observations are set to NaN (since it is not known whether or not it is a malfunction of the measuring instrument) and the simulations are set to the minimum threshold (it would not be physical if the model could not predict at least the lowest measured value of the instruments). However, those situations occur very rarely and this filtering does not significantly impact the analysis.

2) POOL OF LARGE-EDDY SIMULATION MODELS

2.1) REFERENCE VERSUS SENSITIVITY CONFIGURATIONS

A total of eight different LES configurations has been designed and tested for the three LES codes: two configurations for MNH-IBM, three configurations for YALES2, and three configurations for AVBP. Each code is used in its best-known (reference) configuration, i.e. with its a priori best models and options in terms of numerics, boundary conditions or closure models for instance. Since none of these three codes has been used in its classic application framework for the MUST trial, their best-known settings may not hold. So additional test (sensitivity) simulations were carried out to go further into the LES validation. Detailed settings for each reference configuration are given in Table 3. Table 4 highlights the changes between reference and sensitivity simulations.

Table 3: *Reference configurations for MNH-IBM, YALES2 and AVBP for the MUST trial 2681829 (the modified components in the sensitivity configurations are in bold)*

Codes	MNH-IBM	YALES2	AVBP
Code settings			
Navier-Stokes equations	Incompressible, anelastic	Low-Mach	**Artificial compressibility**
Container condition	IBM	Body fitted	Body fitted
Advection scheme	**WENO5 (5th order in space, 4th order in time)**	TFV4A (4th order in space and time)	**TTGC (3rd order in space and time)**
Subgrid turbulence	TKE 1.5	**WALE**	**WALE**
MUST case			
Time step	0.012 s	< 0.02 s	$6.5 \cdot 10^{-4}$ s
Grid type	Structured	**Unstructured**	Unstructured
Grid size	135M cells	**71M cells**	71M cells
Number of cores	1,920	1,600	1,440
Total CPU time	40,000 h	21,000 h	21,600 h

Table 4: *Reference and sensitivity configurations for MNH-IBM, YALES2 and AVBP for the MUST trial 2681829 (the modified components in the sensitivity configurations are in bold)*

Code	LES model name	Advection scheme	Turbulence scheme	Mesh	Compressibility
MNH-IBM	MNH-IBM_reference	WENO5	TKE 1.5	135M cells structured	incompressible
	MNH-IBM_CEN4	**CEN4**	TKE 1.5	135M cells structured	incompressible
YALES2	Y2_reference	TFV4A	WALE	71M cells tetrahedra	low Mach
	Y2_hexa	TFV4A	WALE	**246M cells hexahedra**	low Mach
	Y2_smago	TFV4A	**Smagorinsky**	71M cells tetrahedra	low Mach
AVBP	AVBP_reference	TTGC	WALE	71M cells tetrahedra	artificial
	AVBP_LW	**LW**	WALE	71M cells tetrahedra	artificial
	AVBP_LW_smago	**LW**	**Smagorinsky**	71M cells tetrahedra	artificial

2.2) COMPUTATIONAL GRIDS

The domain of interest is about 225-m by 300-m and includes the 93-m by 171-m container array.

The MUST trial 2681829 was simulated for MNH-IBM over a total computational domain of 300-m long, 300-m wide and 40-m high. The regular Cartesian structured grid is constituted of 135 million hexahedra. The same horizontal and vertical resolutions (0.2 m) are used from the surface to 10 m. Above 10 m, the vertical resolution decreases with a geometrical progression (ratio of 1.08 with the constraint of keeping the vertical resolution below 1 m). Flat ground is boundary-fitted; containers are modelled by an IBM validated by Auguste et al. [1], with at least ten cells per container side to ensure IBM accuracy.

For AVBP and YALES2, the computational domain is meshed with 71 million tetrahedra; the resolution is equal to 0.3 m from the surface until 10 m, and coarsened vertically up to a resolution equal to 1 m at the upper boundary condition. The grid is unstructured and boundary-fitted (Figure 3).

Note that the choice of the computational grid results from a trade-off between accuracy, computational cost and memory issues that is different for each LES code. Note also that there are fundamental geometric differences between MNH-IBM and AVBP/YALES2 computational grids. Changes in the computational mesh have been tested with YALES2: by using hexahedral cells instead of tetrahedral cells in the reference simulation on the one hand, and by increasing the finest mesh resolution from 0.3 m to 0.2 m on the other hand. The objective of these changes was to design a configuration for YALES2 (Y2_hexa in Table 4) that was closer to that of MNH-IBM.

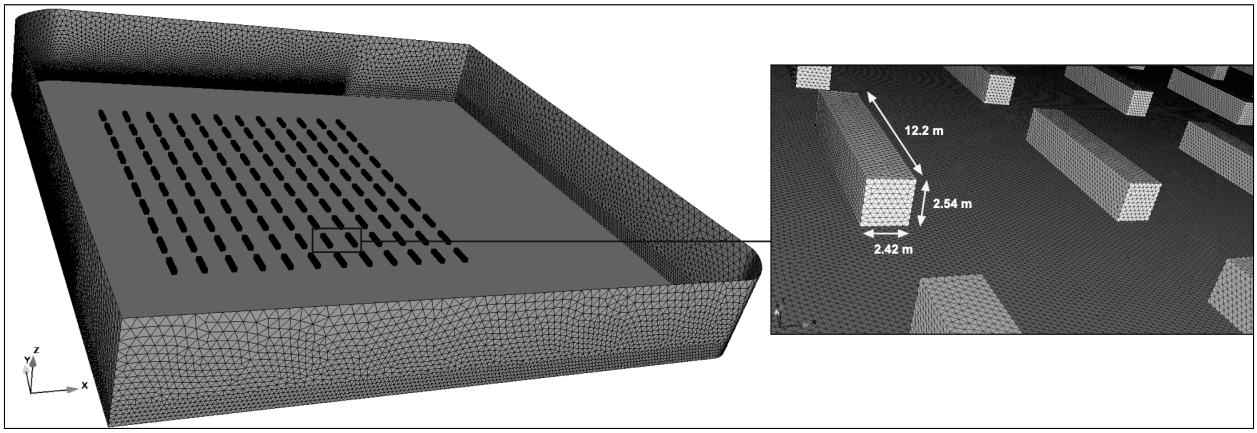


Figure 3: *YALES2 and AVBP unstructured, boundary-fitted mesh (71 million cells, tetrahedra)*

2.3) NUMERICS

a) Advection scheme in MNH-IBM

In its reference configuration, MNH-IBM uses a fourth-order explicit Runge-Kutta scheme for time integration associated with a fifth-order weight essential non-oscillatory (WENO5) scheme [21] for wind advection. WENO5 is used for two main reasons [22]: i) due to its low diffusion properties in the area of sharp gradients, WENO5 features good accuracy and stability with no need of artificial diffusion; and ii) it runs much faster than the fourth-order centered scheme (CEN4) by a factor of 3. The influence of the numerical advection scheme on the MNH-IBM model response is tested, WENO5 being replaced by CEN4 (MNH-IBM_CEN4 in Table 4). In both reference and sensitivity configurations, the Piecewise Parabolic Method (PPM) is used for meteorological variables and tracer advection [7] and the subgrid turbulence model is based on a turbulent kinetic energy (TKE) 1.5 turbulence scheme, where the TKE is a prognostic variable [11].

b) Subgrid turbulence model in YALES2

Equations in YALES2 are discretized with a fourth-order central scheme in space and a fourth-order Runge-Kutta scheme in time (TFV4A) [17]. The subgrid turbulence model for the reference YALES2 simulation is the wall-adaptative local eddy-viscosity (WALE) model [28]. WALE was specifically designed to represent subgrid-scale turbulence in complex geometry, near walls, and to reproduce the proper scaling at the walls.

The influence of the subgrid turbulence model on the YALES2 model response is tested by replacing WALE by the well-known Smagorinsky's model [37] (Y2_smago in Table 4). Note that several LES studies reported in the literature used the Smagorinsky's model [5, 12, 34].

c) Advection scheme and subgrid turbulence model in AVBP

For the reference simulation, AVBP uses a third-order in space and time, explicit, two-step Taylor-Galerkin (TTGC) scheme, which is part of the family of high-order finite-element two-step Taylor-Galerkin/TTG schemes, specifically designed for LES [31, 20]. TTGC is used since it is supposed to be more precise and less dissipative than the standard second-order in space and time Lax-Wendroff (LW) scheme [8]. The subgrid turbulence is modelled by WALE as in the reference YALES2 simulation. The influence of the advection scheme on the AVBP model response is tested as for MNH-IBM, TTGC being replaced by LW (using WALE as in the reference simulation)[2] (AVBP_LW in Table 4). The influence of the subgrid turbulence model on the AVBP model response is tested as for YALES2, WALE being replaced by the well-known Smagorinsky's model [37] (AVBP_LW_smago in Table 4).

2.4) INITIAL AND BOUNDARY CONDITIONS

a) Upstream wind profile

Figure 4 presents the horizontal wind speed data (symbols) available at the upstream tower S at different heights (1.15/4/8/16 m). These data are used to derive initial and inlet wind fields for the LES models.

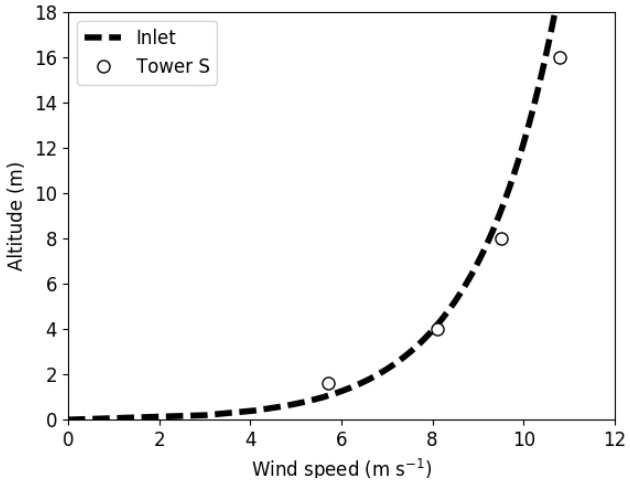


Figure 4: *Vertical profile of the horizontal mean wind speed $\overline{u}(z)$ at the upstream tower S for the MUST trial 2681829. Comparison between experimental data (symbols) and the fitted profile (solid line) obtained using Equation 1 for u_* = 0.715 m s^{-1}.*

The initial and lateral boundary conditions are imposed in the LES models using a mean horizontal upstream wind, which follows a semi-empirical logarithmic law with respect to altitude. Since the thickness of the atmospheric boundary layer over deserts is large and the measured Obukhov length (L_o in Table 1) is high, the vertical profile (z-axis) of the mean horizontal upstream wind $\overline{u}(z)$ is obtained as

$$\overline{u}(z) = \frac{u_*}{\kappa} \ln\left(1 + \frac{z}{z_0}\right) \tag{1}$$

where z_0 (m) is the surface roughness length equal to 0.045 m according to observations [38]; κ is the von Kármán constant equal to 0.4; and u_* (m s^{-1}) is the friction velocity. The formulation of $\overline{u}(z)$ is semi-empirical in the sense that z_0 and u_* are calculated based on horizontal wind speed data available at the upstream tower S at different heights (1.15, 4, 8 and 16 m). The friction velocity is estimated through a least-square regression of Equation 1 to these data. The optimized value of u_* is equal to 0.715 m s^{-1}. The resulting vertical profile (dashed line) is presented in Figure 4 along with the measurements used for regression (symbols).

[2]This configuration is not presented in the 2–D views, data are missing from previous work.

b) Upstream turbulence

Even though the upstream wind is subject to temporal variability (Figure 2), only the mean profile $\overline{u}(z)$ with u_* = 0.715 m s^{-1} and the mean angle $\alpha = -41$ (corresponding to a south-southwest wind) are imposed at the inlet of the computational domain. The reason is two-fold: i) the turbulence generation becomes rapidly controlled by the containers within the canopy, which strongly modifies the near-surface flow topology, and the experimental sensors of interest are located after this transition; ii) the three codes have a very different treatment for turbulence injection, which would introduce a discrepancy detrimental for the comparison of the three codes. Further work could include designing a strategy to inject turbulence and analysing similarities and differences in the LES model response.

c) Ground and container surface

In AVBP and YALES2, the boundary conditions for both ground and containers are a law-of-the-wall based on a viscous length [35]. In MNH-IBM, the ground surface is described with the SURFEX surface scheme [24], which computes the Reynolds stress based on the roughness length, and the container surface is described using the IBM [1].

2.5) PERFORMANCE MEASURES

The present report focuses on the tracer concentration results[3]. They are presented in two ways: i) the plume footprint is given by two-dimensional horizontal views extracted at 1.6-m high (at the same altitude as the probes) and averaging over a 200-s time period; and ii) the statistical metrics by Chang and Hanna [6] are computed for each LES model with respect to the experimental measurements at each of the 25 selected probes (Figure 1).

The statistical metrics compare the tracer concentrations extracted from the simulations with those measured experimentally in terms of normalized mean square error (NMSE), fractional bias (FB), fraction of predictions within a factor of two of observations (FAC2), geometric mean bias (MG) and geometric variance (VG):

$$\mathrm{NMSE} = \frac{\overline{(C_o - C_s)^2}}{\overline{C_o}\,\overline{C_s}}, \tag{2}$$

$$\mathrm{FB} = \frac{(\overline{C_o} - \overline{C_s})}{0.5\,(\overline{C_o} + \overline{C_s})}, \tag{3}$$

$$\mathrm{FAC2} = \text{fraction of data that satisfies } 0.5 \leq \frac{C_s}{C_o} \leq 2.0, \tag{4}$$

$$\mathrm{MG} = \exp\left(\overline{\ln C_o} - \overline{\ln C_s}\right), \tag{5}$$

$$\mathrm{VG} = \exp\left[\overline{(\ln C_o - \ln C_s)^2}\right], \tag{6}$$

where C_o and C_s are the observed and simulated concentrations, and the overbar denotes the mean value over the dataset. FB and MG measure the mean relative bias and are an indicator of systematic errors. NMSE and VG measure the mean relative scatter and indicate both systematic and random errors. A perfect model would have NMSE and FB equal to 0; FAC2, MG and VG equal to 1.

2.6) AVBP SPECIFIC FEATURES

a) Artificial compressibility

Since AVBP is a fully compressible LES solver, the time-step is restricted by the Courant-Friedrichs-Lewy (CFL) condition for the acoustic wave propagation. Since correctly capturing the acoustic waves is not relevant in the context of atmospheric boundary layer simulation (very low Mach number), an artificial compressibility approach [32], also known as pressure gradient scaling (PGS) and referred to as “cryogenic”, is used for all AVBP configurations in the pool of LES models to increase the time step and gain computational time.

[3] Rochoux et al. [33] present wind speed profiles at the central tower T in addition to the tracer concentration results.

The key idea of artificial compressibility is to artificially decrease the propagation speed of the acoustic waves (c) by rescaling the pressure p and the temperature T by a factor α^2, leading to artificial temperature $T^* = \alpha^2 T$ and pressure a$p^* = \alpha^2 p$. With this transformation, the resulting acoustic wave propagation speed (c^*) is artificially reduced by a factor α:

$$c^* = \sqrt{\gamma r T^*} = \sqrt{\gamma r \alpha^2 T} = \alpha c \tag{7}$$

where γ is the heat capacity ratio, which significantly reduces the constraint on the AVBP model time-step. It can be shown that the hydrodynamic problem remains unchanged [32], provided the resulting artificial Mach number $\mathrm{Ma}^* = \mathrm{Ma}/\alpha$ remains small. Recall that the Mach number is defined as $\mathrm{Ma}^* = \frac{u}{c^*} = \frac{u}{\alpha c}$, where u is the mean upstream wind speed. In the present study, the magnitude of the incoming wind speed is about 6.4 m s^{-1} at 1.6 m high. α is thereby set to 0.22 to guarantee that $\mathrm{Ma}^* < 0.1$ ($\mathrm{Ma}^* = 0.086$ at 1.6 m high) in the whole simulation domain. The new CFL condition satisfies $\mathrm{CFL} = (u + c^*)\,\frac{\Delta t^*}{\Delta x}$, where Δt is the model time step and Δx is the mesh step size. Since the acoustic wave propagation speed is reduced by a factor α, the time step Δt^* can be increased without changing the CFL condition.

We check the quality of the artificial compressibility solution (AVBP_reference in Table 5) with the reference configuration considering the full compressible solver in AVBP (AVBP_no_cryo). Figure 5 compares the (time-averaged) mean plume footprint and Table 5 compares statistical metrics for both configurations.

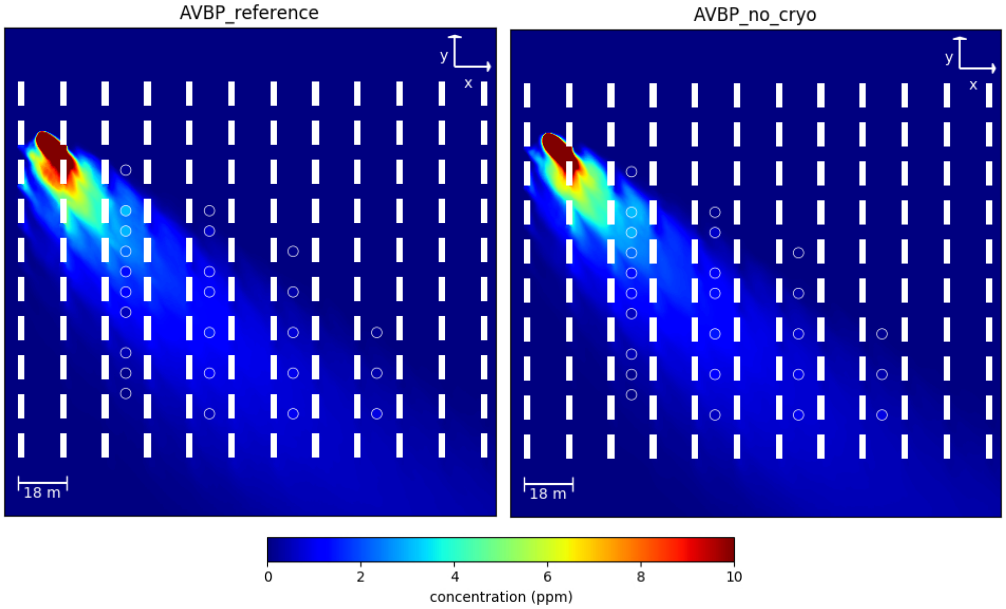


Figure 5: *Mean horizontal tracer concentration (ppm) at 1.6-m high obtained with AVBP configurations, with (AVBP_reference, left panel) and without (AVBP_no_cryo, right panel) artificial compressibility. White rectangles represent containers. Probe positions are indicated by circle symbols. They are colored using experimental data using the same color map as for simulated data. Values are truncated at 10 ppm for image clarity purpose.*

Field results show that plume footprints are similar between both AVBP configurations. We can still notice some small differences in the lateral spreading of the plume, there are some tracer accumulation upwind the first container encountered by the tracer in the case with artificial compressibility (left panel in Figure 5) that is not as significant in the case without artificial compressibility (right panel in Figure 5). Consistently, statistical results are very close to each other. The NMSE is small for both configurations, the FB has the same negative sign (i.e. the present AVBP models tend to overestimate observations) and the FAC2 is reduced by a few % when using artificial compressibility (the FAC2 is upgraded from 60% to 56%). The computational cost is significantly reduced (by a factor of about 5, changing the total CPU time from 120,000 hours to 21,600 hours for typical HPC resources from GENCI), making AVBP competitive with MNH-IBM and YALES2 (Table 3). These elements motivate the use of artificial compressibility for all AVBP configurations in the present pool of LES models (Table 4).

Table 5: *Comparison of AVBP reference solutions with and without artificial compressibility in terms of standard statistical metrics*

LES model name	NMSE	FB	FAC2	MG	VG
AVBP_reference (with)	0.0065	-0.08	0.56	1.22	2.18
AVBP_no_cryo (without)	0.0231	-0.15	0.60	1.13	2.08

b) Treatment of boundary conditions

Characteristics-based NSCBC (Navier-Stokes Characteristics Boundary Conditions) are used at inlet and outlet in AVBP to properly handle acoustics [15, 29]. This is required even when using artificial compressibility. Relaxation coefficients are therefore introduced to avoid numerical instabilities and spurious acoustic wave reflections at the computational domain boundaries. We give here some additional, technical details about the AVBP numerical configuration. This part might not be easily understandable for a non-AVBP user.

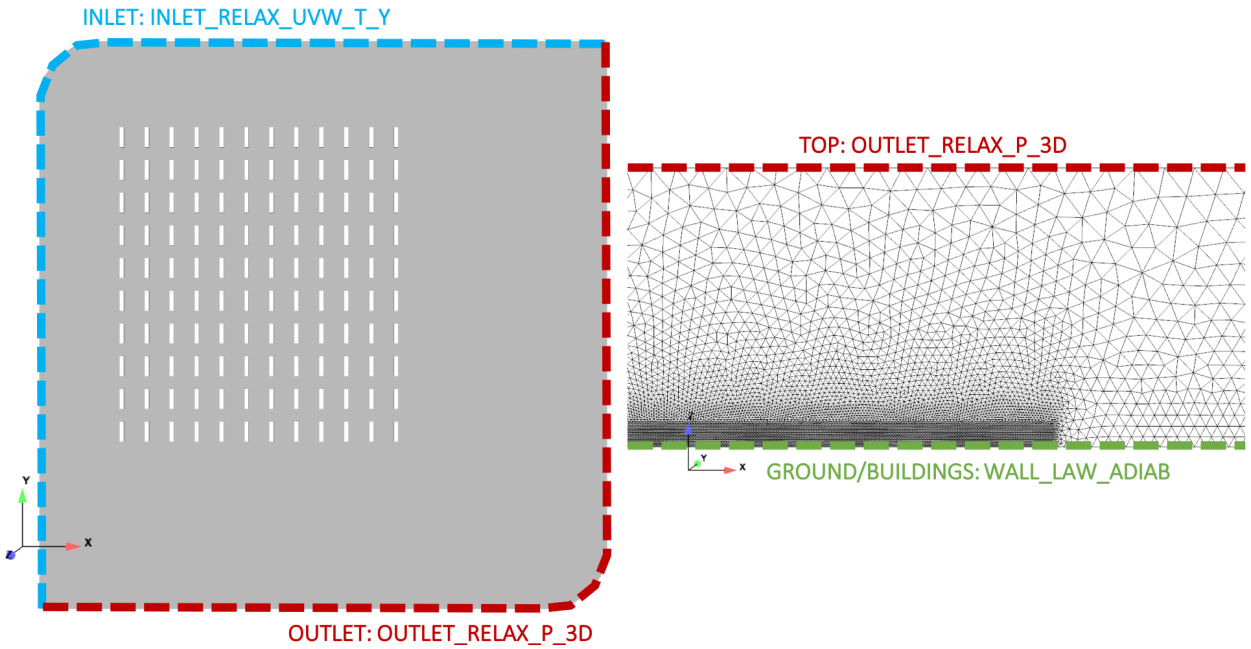


Figure 6: *Schematics of AVBP boundary conditions: inlet, outlet, top and (ground/building) surface.*

In spite of the NSCBC boundary conditions, edge effects occurred in the AVBP reference configuration. Typically, a large non-physical increase in pressure and velocity appears near the outlet boundary. To overcome this issue, two approaches were tested: i) modifying the relaxation coefficient (from 1 to 0.1) on the outlet patch of the domain (AVBP_RELAX), and ii) adding a sponge layer near the outlet patch with increased artificial viscosity (to 0.2 $\mathrm{kg\,m^{-1}\,s^{-1}}$) (AVBP_VISCOBOUND). Figure 7 shows that the relaxation coefficient does not act on the numerical solution instability as evidenced by the edge effects on pressure, temperature, wind speed and artificial viscosity fields (left panels). The AVBP_VISCOBOUND configuration is not subject to these edge effects and thereby appears as the best solution to overcome this numerical issue. Hence, all AVBP configurations presented in Table 4 are carried out using the VISCOBOUND approach.

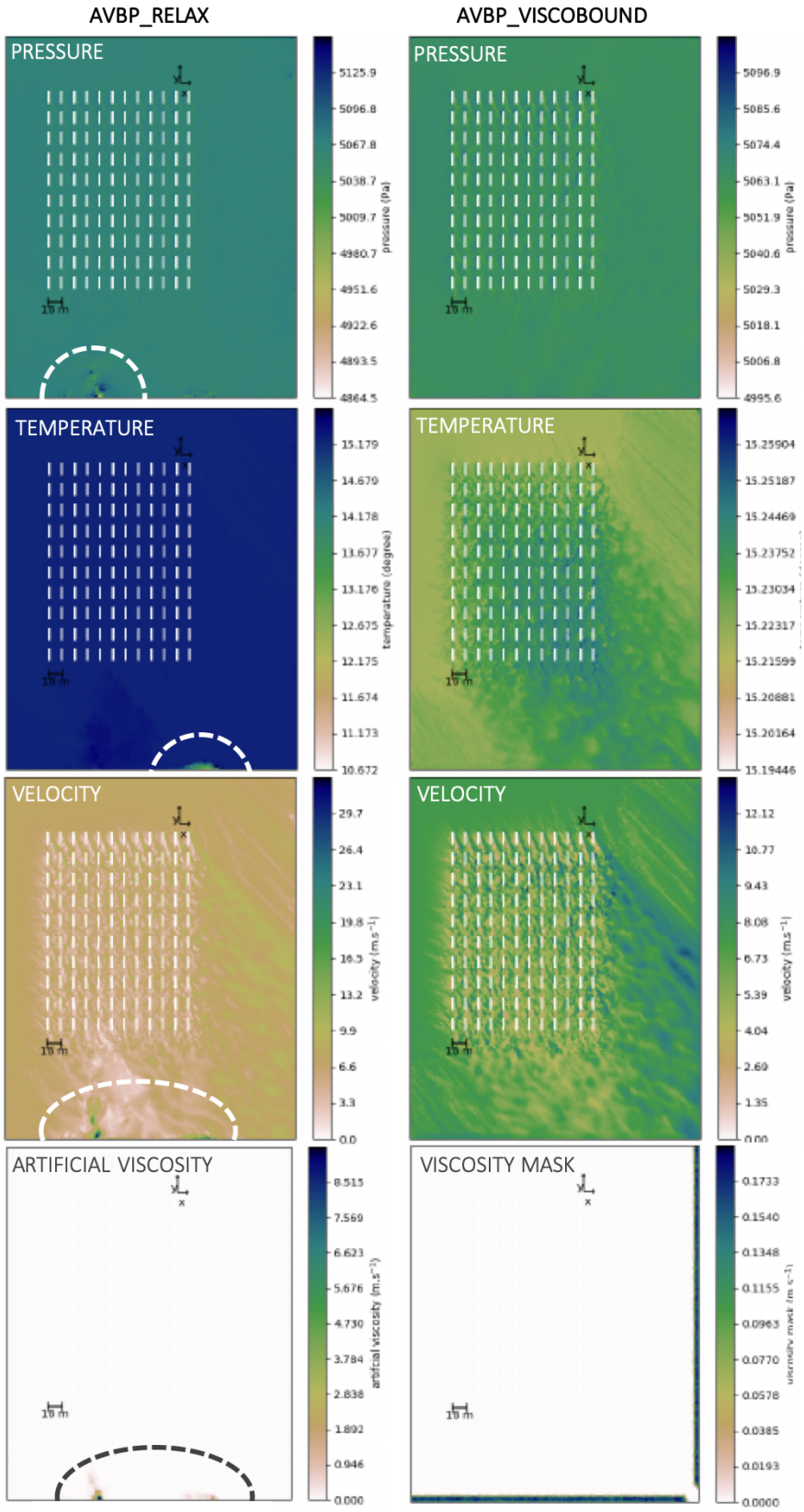


Figure 7: *Mean horizontal pressure, temperature, velocity magnitude and viscosity at 1.6 m high from top to bottom panels for AVBP_RELAX (left panels) and AVBP_VISCOBOUND (right panels) configurations. For the viscosity panels, the left panel represents artificial viscosity generated by the LES code and the right panel represents the viscosity mask (or sponge layer) defined a priori by the user. White rectangles represent containers. Dashed lines indicate the edge effect areas.*

Table 6: *Features of the inlet/lateral boundary condition of type INLET_RELAX_UVW_T_Y (Figure 6), where the profile of the streamwise flow component (U) is given by the logarithmic law in Equation 4 (Figure 1)*

patch_name	INLET
boundary_condition	INLET_RELAX_UVW_T_Y
target_origin	solutbound
wave_type	temporal
relax_type	local
relax_coeff_Un	1.0D0
relax_coeff_Ut	2.0D0
relax_coeff_T	1.0D0
relax_coeff_Y	1.0D0
U	*profile*
V	0.0d0
W	0.0d0
Temperature	304.4d0
P	1.01300d5
AIR	1.0d0
tracer	0.0d0

Table 7: *Features of the outlet and top boundary conditions of type OUTLET_RELAX_P_3D (Figure 6)*

patch_name	OUTLET/TOP
boundary_condition	OUTLET_RELAX_P_3D
target_origin	solutbound
wave_type	normal_FE
relax_type	local
relax_coeff_P	1D0
damping_type	local
pressure	1.01300d5

Table 8: *Features of the ground and building boundary conditions of type WALL_LAW_ADIAB (Figure 6)*

patch_name	GROUND/BUILDINGS
boundary_condition	WALL_LAW_ADIAB

3) MODEL-TO-MODEL COMPARISON RESULTS

We propose both a qualitative analysis of spatially-distributed tracer dispersion and a statistical analysis for each LES configuration to provide an exhaustive model-to-model comparison. All tested configurations with MNH-IBM, YALES2 and AVBP are presented in Table 4.

3.1) STATISTICAL ANALYSIS

a) Comparison of statistical metrics

Table 9 reports the standard statistical metrics by Chang and Hanna [6] that are computed over the set of 25 selected probes for the different configurations simulated with MNH-IBM, YALES2 and AVBP. Since the FAC2 is the most robust metric for air quality performance evaluation [16], we mainly focus the analysis in Rochoux et al. [33] on the FAC2 score. The FB score is also used to evaluate the simulation-observation discrepancies (normalized by the average of the respective simulation and observation). The FB score thereby indicates if a given LES configuration overpredicts the measurements (FB < 0) or underpredicts them (FB > 0). For completeness, the present report provides all the statistics for all configurations. In complement, Table 10 gives the tracer concentration statistics (mean, standard deviation, 99th percentile, maximum) obtained over the same set of probes.

Table 9: *Standard statistical metrics obtained for MNH-IBM, YALES2 and AVBP configurations. In bold is indicated the best configuration for each LES code, "best" meaning having the highest FAC2.*

LES model	NMSE	FB	FAC2	MG	VG
MNH–IBM_reference	0.1162	-0.34	**0.64**	1.25	1.95
MNH–IBM_CEN4	0.0704	-0.27	0.60	1.44	2.58
Y2_reference	0.0123	-0.11	0.76	1.24	1.68
Y2_hexa	**0.0091**	**-0.10**	**0.76**	1.30	1.62
Y2_smago	0.0145	0.12	0.76	1.41	1.73
AVBP_reference	0.0065	-0.08	0.56	1.22	2.18
AVBP_LW	0.0021	0.05	**0.84**	1.33	1.51
AVBP_LW_smago	0.0004	-0.02	0.72	1.15	1.59

Results show that all LES models are in the acceptable range in terms of FAC2 (above 50%) and feature a small NMSE. Most of them have a negative FB and thereby tend to overestimate observations. This is consistent with the overestimation of tracer concentration statistics in Table 10. MG values are at the acceptability threshold (fixed to 1.3 according to Chang and Hanna [6]) and VG values are above the acceptability threshold (fixed to 1.6) except for AVBP_LW and AVBP_LW_smago.

The best configuration for MNH-IBM is the reference one corresponding to the WENO5 numerical scheme with FAC2 equal to 64%. Similar tracer concentration statistics are obtained for both MNH-IBM configurations; they are overestimated by about 40%. In contrast, AVBP and YALES2 achieve their best performance in the sensitivity tests for different configurations. For YALES2, the three tested configurations feature the same FAC2 score of 76%. By considering also NMSE, FB and tracer concentration statistics, Y2_hexa can be considered as the best configuration. In particular, Y2_hexa provides very good approximation of tracer concentration statistics (e.g. mean overestimation by 10%, perfect estimation of standard deviation). Still, the three YALES2 configurations remain in a very narrow range. There is much more variability among the AVBP configurations: the FAC2 score increases from 56% in AVBP_reference to 84% in AVBP_LW. Based on the FAC2 score, AVBP_LW appears as the best set-up among the pool of eight LES models to simulate the present MUST trial. AVBP_LW features a low NMSE, a low FB (the FB is positive, meaning that this configuration tends to underestimate observations) and good tracer concentration mean and standard deviation approximation.

Figure 8 presents scatter plots of time-averaged model predictions for all 25 probes (1 dot corresponds to 1 probe and 1 color corresponds to 1 LES code). Right panels correspond to reference configurations; left panels correspond to the best configurations. The Pearson correlation coefficient (R) is indicated. All configurations feature a correlation coefficient above 0.85. Some changes are visible for the YALES2

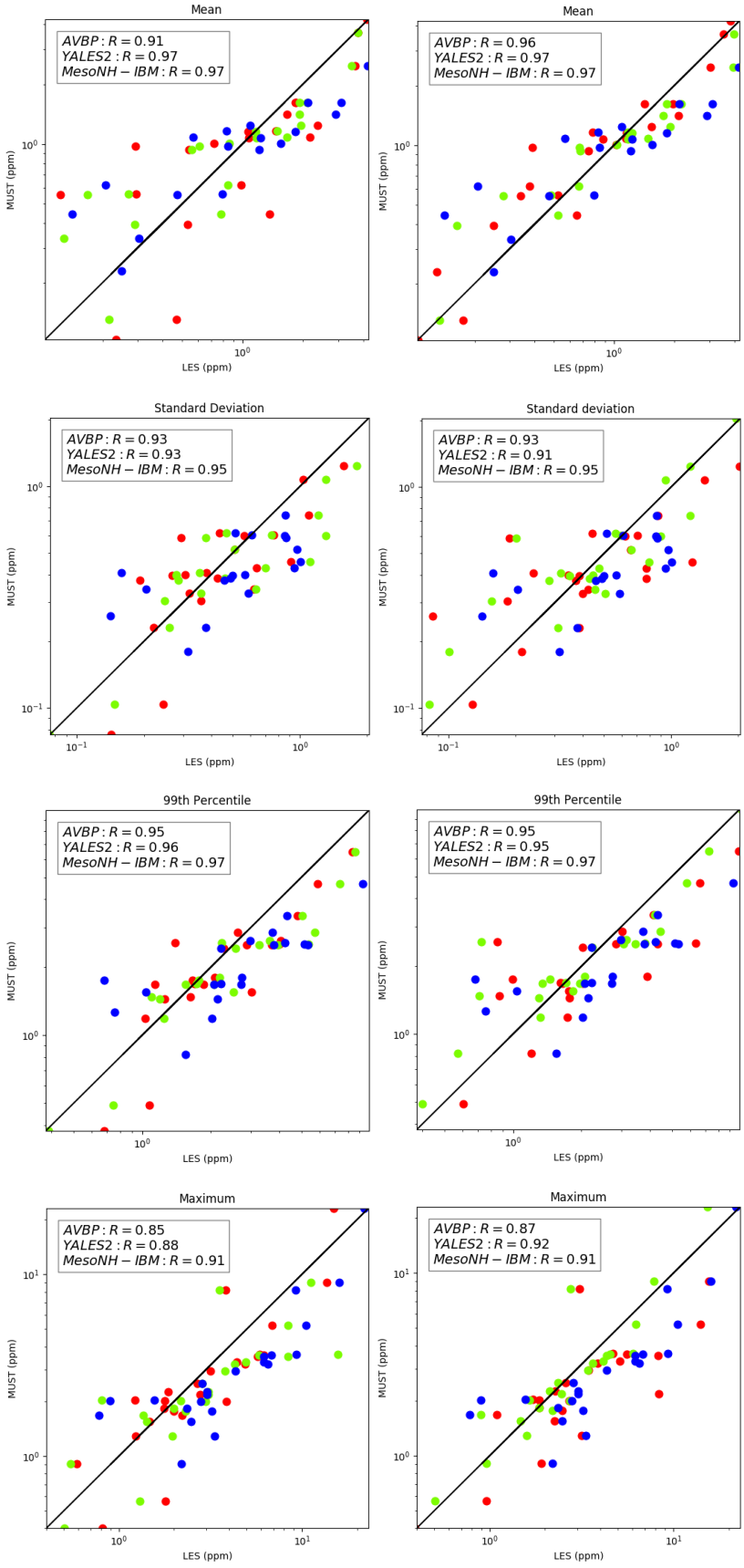


Figure 8: *Scatter plots of measured (x-axis) versus simulated (y-axis) mean, standard deviation, 99th percentile and maximum tracer concentration at each sensor for the three LES codes (Meso-NH, YALES2 and AVBP) in the reference configurations (left panels) and in the best configurations (right panels). The correlation coefficient R is indicated.*

Table 10: *Comparison of tracer concentration statistics (in ppm) between observations and simulations obtained at the 25 sensors. The relative error with respect to the observation (in %) is indicated in brackets.*

LES model	Mean	Standard deviation	99th percentile	Maximum
MNH–IBM_reference	1.54 (+40%)	0.77 (+48%)	3.57 (+47%)	5.04 (+40%)
MNH–IBM_CEN4	1.43 (+30%)	0.68 (+31%)	3.27 (+35%)	4.28 (+19%)
Y2_reference	1.23 (+12%)	0.63 (+21%)	3.12 (+28%)	4.80 (+33%)
Y2_hexa	1.21 (+10%)	0.52 (0%)	2.53 (+4%)	3.19 (-11%)
Y2_smago	0.97 (-12%)	0.60 (+15%)	3.03 (+25%)	5.28 (+47%)
AVBP_reference	1.23 (+12%)	0.54 (+4%)	2.72 (+12%)	3.83 (+6%)
AVBP_LW	1.05 (-5%)	0.64 (+23%)	3.06 (+26%)	4.85 (+35%)
AVBP_LW_smago	1.12 (+2%)	0.67 (+29%)	3.23 (+33%)	5.74 (+59%)
Observation	1.10	0.52	2.43	3.60

configurations with a significant improvement in the maximum concentrations for Y2_hexa (R changing from 0.88 to 0.92). The mean concentrations are also significantly improved when changing to AVBP_LW (R changing from 0.91 to 0.96). This provides some insights into the best configurations of AVBP and YALES2.

b) Local versus background statistical metrics

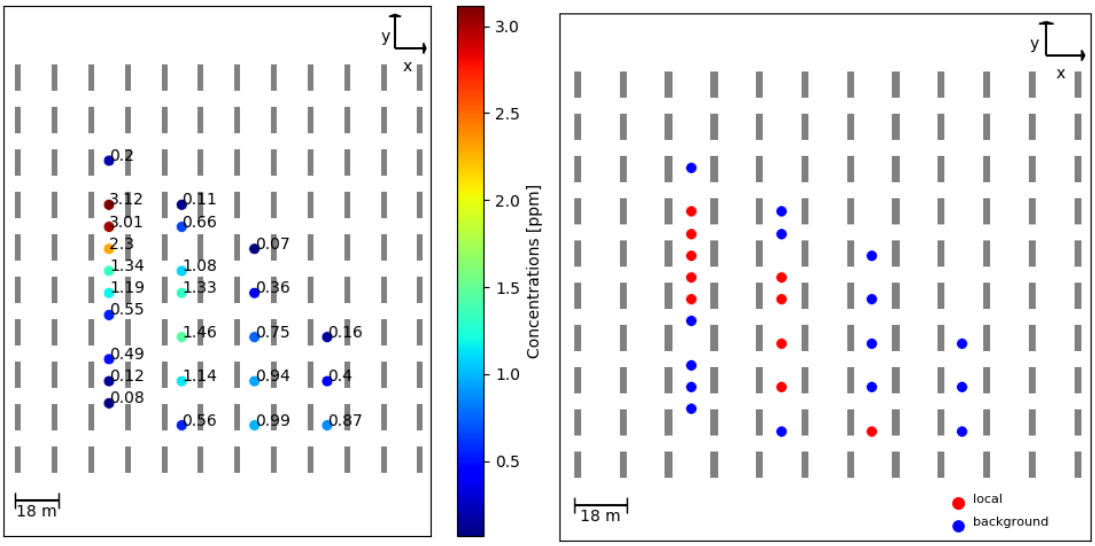


Figure 9: *Clustering of local and background tracers among the 25 probes. Left panel: map of the time-averaged observed tracer concentration for the MUST trial 2681829 (in ppm). Right panel: corresponding classification of local (red) and background (blue) probes.*

To go further in the analysis, we cluster the model performance by classifying tracer concentrations into two categories. We separate local concentrations corresponding to near-source high concentrations above 1 ppm from background concentrations corresponding to low concentrations below 1 ppm further away from the plume hot spot. This clustering between background and local concentrations is determined based on the observed data at each probe; left panel in Figure 9 shows the 200-s time-averaged observed concentration at each probe. Consistently, right panel in Figure 9 indicates the label for each probe:
• local probes (red symbols): 10 probes numbered 5, 6, 7, 8, 9, 14, 15, 16, 17, 22;
• background probes (blue symbols): 15 probes numbered 1, 2, 3, 4, 10, 13, 18, 19, 22, 23, 24, 25, 26, 31, 32, 33.

Table 11 presents the statistical metrics obtained for local and background probes for MNH-IBM, YALES2 and AVBP, respectively. Local concentrations are better simulated with AVBP and YALES; the local FAC2 is equal to 100%, except for AVBP_reference (80%). MNH-IBM exhibits a local FAC2 ranging from 70 to 80%. Background concentrations are not as well represented as local concentrations, especially for

AVBP_reference for which the FAC2 is below the acceptable value (40%). MNH-IBM, YALES2 and AVBP_LW_Smago configurations are slightly better with a FAC2 ranging from 53 to 60%. But what is noticeable is that the AVBP_LW configuration provides the best representation of background tracer concentration with a background FAC2 equal to 73%.

Table 11: *Local and background statistical metrics obtained for MNH-IBM, YALES2 and AVBP*

MNH-IBM	**_reference**		**_CEN4**	
	local	background	local	background
NMSE	0.20643	0.00003	0.11672	0.00238
FB	0.440	0.005	0.337	0.049
FAC2	0.80	0.53	0.70	0.53
MG	0.78	1.72	0.96	1.89
VG	1.30	2.55	1.50	3.70
Mean sim. (ppm)	3.07 (+57%)	0.52 (-2%)	2.75 (+40%)	0.55 (+4%)
Mean obs. (ppm)	1.96	0.53	1.96	0.53

Y2	**_reference**		**_hexa**		**_smago**	
	local	background	local	background	local	background
NMSE	0.05215	0.06621	0.04323	0.06399	0.00255	0.10159
FB	0.227	0.255	0.207	0.251	0.050	0.315
FAC2	1.00	0.60	1.00	0.60	1.00	0.60
MG	0.78	1.70	0.81	1.79	0.99	1.78
VG	1.08	2.25	1.07	2.14	1.06	2.41
Mean sim. (ppm)	2.46 (+26%)	0.41 (-23%)	2.41 (+23%)	0.41 (-23%)	1.86 (-5%)	0.38 (-28%)
Mean obs. (ppm)	1.96	0.53	1.96	0.53	1.96	0.53

AVBP	**_reference**		**_LW**		**_LW_smago**	
	local	background	local	background	local	background
NMSE	0.02888	0.03347	0.00114	0.07517	0.00635	0.01960
FB	0.169	0.182	0.034	0.272	-0.080	0.140
FAC2	0.80	0.40	1.00	0.73	1.00	0.53
MG	0.82	1.58	0.98	1.63	0.91	1.35
VG	1.14	3.36	1.05	1.92	1.05	2.10
Mean sim. (ppm)	2.32 (+18%)	0.44 (-17%)	2.02 (+3%)	0.40 (-25%)	2.12 (+8%)	0.46 (-13%)
Mean obs. (ppm)	1.96	0.53	1.96	0.53	1.96	0.53

The results obtained for local and background simulations are satisfactory for all configurations. AVBP is the model with the highest sensitivity to its configuration since it can be both the best (AVBP_LW) and the least accurate (AVBP_reference) configuration among the pool of LES models. The AVBP_LW configuration obtains the best FAC2 results due to its improved representation of the background concentration. When using the Smagorinsky subgrid turbulence model (AVBP_LW_Smago), the background FAC2 is degraded to 53%, while the local FAC2 remains equal to 100%. AVBP results tend to show that the advection scheme impacts both local and background concentration statistics, but the subgrid turbulence model mostly impacts background concentration statistics. In contrast, for MNH-IBM and YALES2, scores are in a very narrow range in-between the configurations. Hence, MNH-IBM and YALES do not show any particular sensitivity to changes in their configurations for the given boundary conditions.

c) Impact of instrument detection threshold

Milliez and Carissimo [26] stated that the CFD model performance could be improved by considering a detection threshold of 0.1 ppm (instead of 0.04 ppm) to filter out very low tracer concentration values. From an experimental viewpoint, it is difficult to know if these low values are the result of a physical measurement or if they result from sensor malfunction. From a CFD viewpoint, models generate numerical errors and can provide irrelevant model outputs, especially for very low tracer concentration values that can often be considered as noise. Table 12 presents the statistical metrics obtained for each LES model as a function of the

detection threshold. This is only done for the reference configurations to test if this change in concentration threshold can improve our LES results.

Table 12: *Statistical metrics obtained for MNH–IBM, YALES2 and AVBP reference configurations according to the instrument detection threshold (0.04 versus 0.1 ppm)*

	MNH-IBM_reference		**Y2_reference**		**AVBP_reference**	
	0.04 ppm	0.1 ppm	0.04 ppm	0.1 ppm	0.04 ppm	0.1 ppm
NMSE	0.1162	0.1010	0.0123	0.0100	0.0065	0.0053
FB	-0.34	-0.33	-0.11	-0.10	-0.08	-0.07
FAC2	0.64	0.72	0.76	0.80	0.56	0.60
MG	1.25	1.14	1.24	1.18	1.22	1.15
VG	1.95	1.51	1.68	1.38	2.18	1.62

The change in detection threshold significantly improves model performance: a gain in FAC2 of a few % (ranging from 4% for YALES2 and AVBP to 8% for MNH-IBM) is obtained; MG and VG are now within the acceptability limits as defined by Hanna, Hansen, and Dharmavaram [16]. A larger detection threshold benefits to MNH–IBM, which tends to underestimate low values on the plume lateral sides. Using the 0.1-ppm detection threshold reduces the weight of very low tracer concentration values, which are observed far away from the plume centerline.

3.2) QUALITATIVE COMPARISON

Figure 10 represents the horizontal dispersion of the plume at 1.6 m high for some of the LES models presented in Table 4. Left panels present the reference configurations for the three codes (MNH-IBM, YALES2 and AVBP); right panels present one sensitivity configuration per code: MNH-IBM_CEN4, Y2_hexa and AVBP_LW_Smago. This does not correspond to the AVBP best configuration: some data are missing from this configuration preventing us from plotting the time-averaged tracer concentration field.

Two types of plume shape can easily be identified based on the two-dimensional tracer dispersion maps presented in Figure 10: i) stretched plumes in the inlet wind direction (MNH-IBM_reference, MNH-IBM_CEN4, Y2_hexa); and ii) short compact plumes that are more dispersed laterally (Y2_reference, AVBP_reference, AVBP_LW_Smago).

a) Stretched plumes

Stretched plumes are defined by a very narrow plume advected along the same direction. Lateral dispersion (in a direction perpendicular to the plume centerline) is weak, suggesting that advection dominates turbulence. We observe a division of the plume into two parts around the first container met by the tracer. An accumulation zone is formed upstream of the container, while much of the tracer is advected into the open area in-between the containers acting as a channel for the flow. Consistently, wind roses presented in Figure 11 show that the wind direction within the container array mainly follows the south-southwest inlet wind direction and is not significantly deflected by the presence of obstacles. Moderate wind conditions are observed. The magnitude of the incoming wind speed is about 6.4 $\mathrm{m\,s^{-1}}$ at 1.6 m high and more than 50% of the simulated wind within the container array is more than 5 $\mathrm{m\,s^{-1}}$. Such moderate wind conditions can significantly contribute to reduce the plume lateral dispersion, as it is the case for MNH-IBM configurations and also Y2_hexa.

b) Short compact plumes

Short compact plumes include Y2_reference, AVBP_reference and AVBP_LW_Smago configurations, which share the same computational grid. These configurations feature a similar plume geometry with a very localised area of high tracer concentration upstream the first containers met by the tracer and a change of the plume centreline direction consistently with actual observations of the MUST trial. As shown by wind roses in Figure 11, there is a stronger impact of the containers on the flow dynamics than with MNH–IBM or Y2_hexa. The wind flow is mainly oriented in the vertical direction of the containers (about 40 to 50%, especially for YALES2), with slightly faster wind speeds for YALES2 (ranging from 4 to 6 $\mathrm{m\,s^{-1}}$) than for

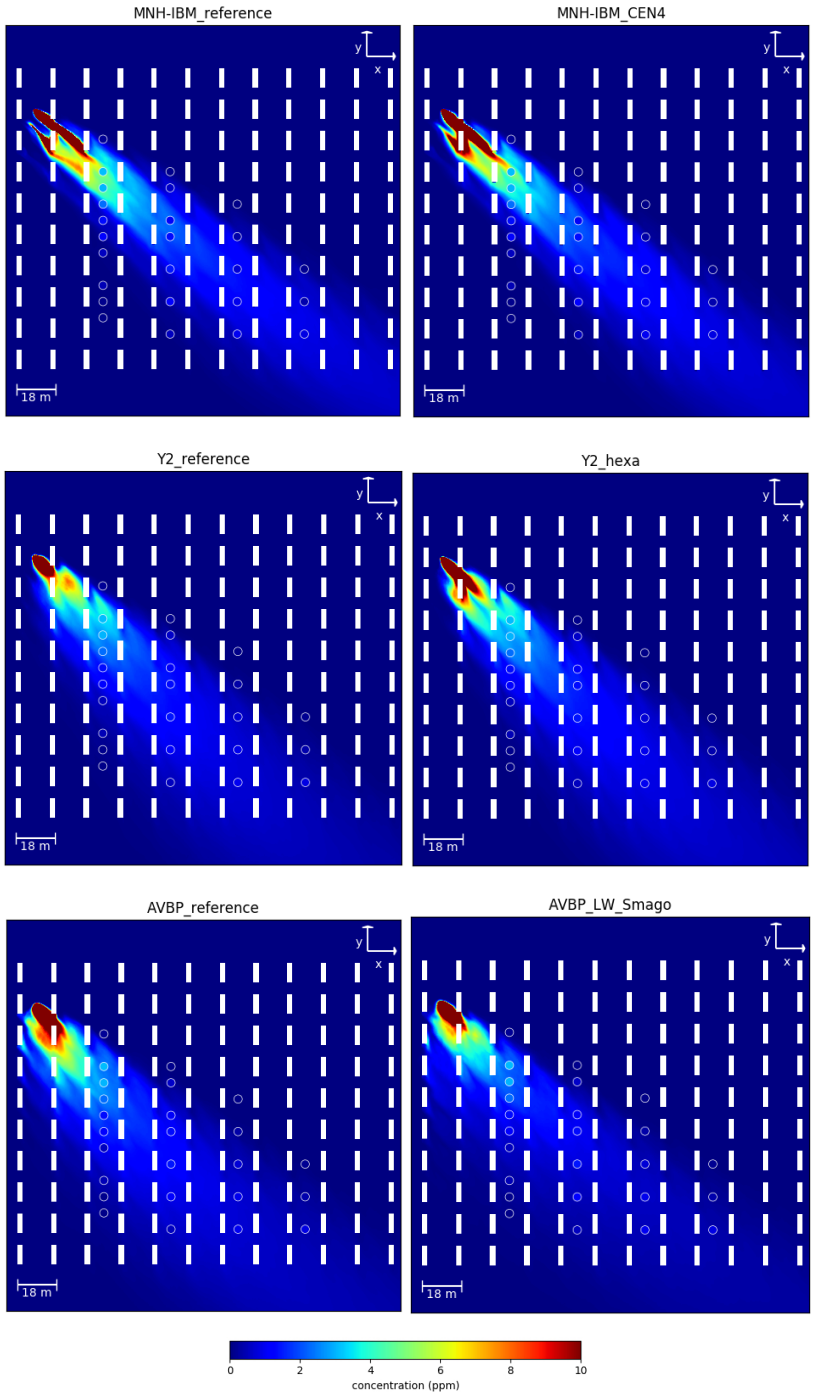


Figure 10: *Mean horizontal tracer concentration (ppm) at 1.6 m high for MNH-IBM, YALES2 and AVBP configurations (from top to bottom panels). Left panels: reference simulations (MNH-IBM_reference, Y2_reference and AVBP_reference). Right panels: sensitivity simulations (MNH-IBM_CEN4, Y2_hexa and AVBP_LW_smago). The observed mean concentration (ppm) at the sensor locations is also given by the symbol colors (circles) using the same color map, which only represents values from 0 to 10 ppm (very strong values in the vicinity of the source term are not shown for clarity purpose). White rectangles represent containers.*

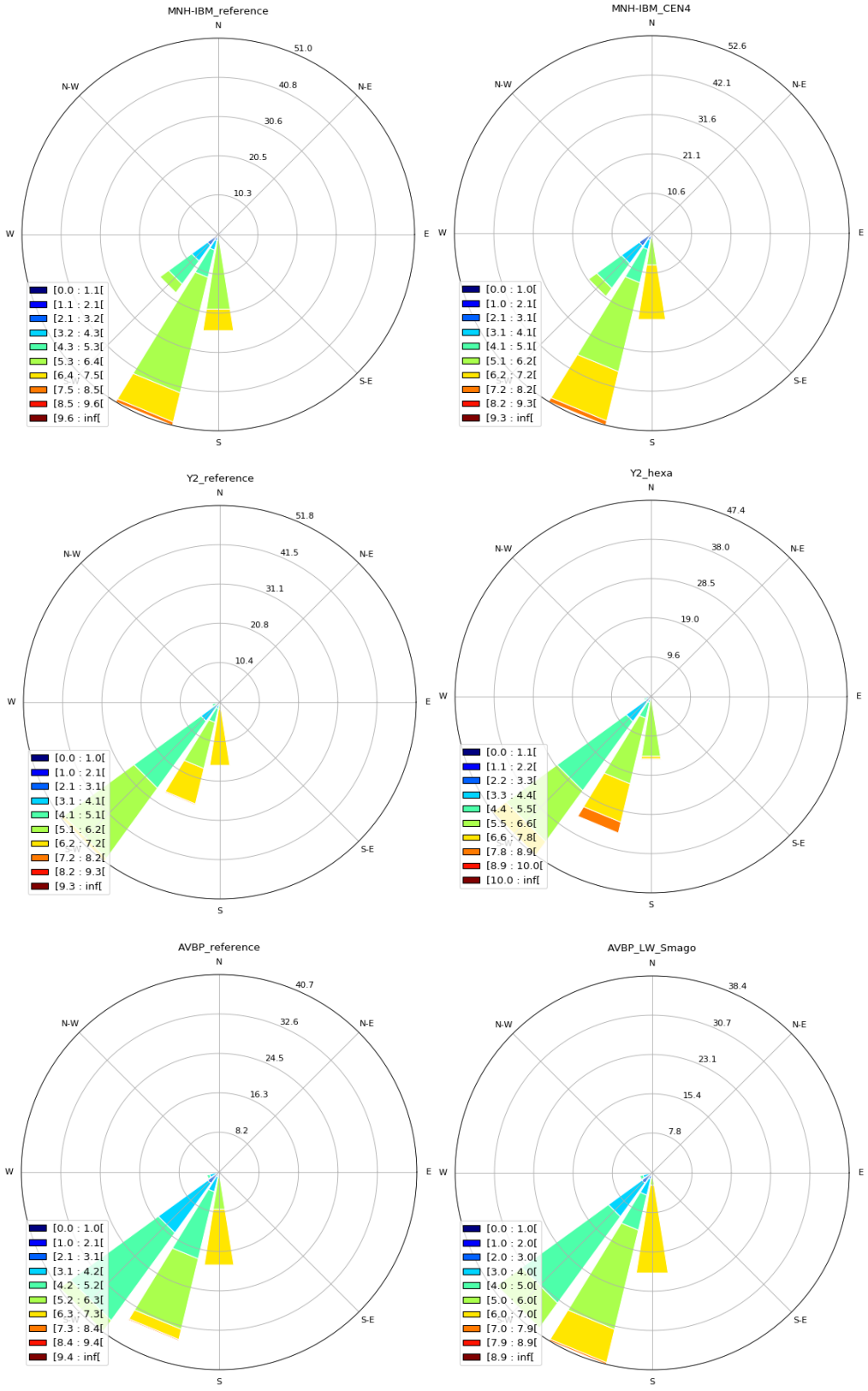


Figure 11: *Wind roses representing the percentage frequency distribution of wind speed ($m\,s^{-1}$) and direction (°C) at 1.6 m high fMNH-IBM, YALES2 and AVBP configurations (from top to bottom panels) corresponding to Figure 10. The percentage frequency is indicated with the grey circles: the larger the circle radius, the larger the percentage frequency. The colors represents the wind speed that is divided into ten categories.*

AVBP (ranging from 4 to 5 $\mathrm{m\,s^{-1}}$). The resolution of the mesh and the type of mesh cells might influence the dispersion since Y2_hexa is in the first category of stretched plumes and the only difference with respect to Y2_reference is the computational grid. Note that dispersion is expected to increase with mesh size. It would be of interest to study if AVBP features the same trend as YALES2 when changing the computational mesh. Note also that the standalone effect of the LW scheme in the AVBP model response cannot be analysed here in terms of plume shape since data are missing.

4) STRUCTURAL MODEL UNCERTAINTY ESTIMATION APPROACH

Testing multiple models in different configurations generates a multi-model ensemble (MME) or a pool of LES models, which samples the structural model uncertainty. Statistics can be derived to evaluate the variability among the pool of LES results using observations as a reference. Specifically, we are interested in the minima and maxima of our MME to obtain the "envelope" of the LES models.

a) Definition of multi-model ensemble (MME) error and variability

To evaluate the MME error with respect to observations, we use the maximum error between simulations and observations characterized by its (time-averaged) mean value $\overline{E}_{max}$ and its standard deviation σ_{Emax}:

$$\overline{E}_{max} = \max_{c} \left(\frac{1}{n_t} \sum_{t=1}^{n_t} \left| x_{s,c}^{t} - x_{o}^{t} \right| \right) \tag{8}$$

$$\sigma_{Emax} = \max_{c} \left(\sqrt{\frac{1}{n_t - 1} \sum_{t=1}^{n_t} \left(x_{s,c}^{t} - x_{o}^{t} - \frac{1}{n_t} \sum_{t=1}^{n_t} \left| x_{s,c}^{t} - x_{o}^{t} \right| \right)^2} \right) \tag{9}$$

where x corresponds to the tracer concentration at a given time index t over n_t time steps (i.e. over the 200-s time sequence), o corresponds to the observation, s corresponds to the simulation and c corresponds to the LES model configuration.

To evaluate the MME variability, we use the difference between model maxima and minima characterized by its (time-averaged) mean value $\overline{\Delta}$ and its standard deviation σ_{Δ}.

$$\overline{\Delta} = \max_{c} \left(\frac{1}{n_t} \sum_{t=1}^{n_t} x_{s,c}^{t} \right) - \min_{c} \left(\frac{1}{n_t} \sum_{t=1}^{n_t} x_{s,c}^{t} \right) \tag{10}$$

$$\sigma_{\Delta} = \frac{1}{n_c - 1} \sum_{c=1}^{n_c - 1} \left(\frac{1}{n_t} \sum_{t=1}^{n_t} x_{s,c}^{t} - \frac{1}{n_t \cdot c} \sum_{c=1}^{n_c} \sum_{t=1}^{n_t} x_{s,c}^{t} \right)^2 \tag{11}$$

where n_c corresponds to the number of LES configurations. The MME variability indicates the degree of confidence we can have in LES models for near-field dispersion problems but can also help spotting the configurations that are far from the MME mean and that can be considered as outlier.

b) Envelopes and statistics of multi-model ensemble (MME)

To visualize the MME error and the MME variability, we select only three configurations (i.e. one per LES code). A first set of configurations corresponds to the reference LES models (MNH-IBM_reference, Y2_reference, and AVBP_reference). A second set of configurations corresponds to the best configurations (MNH-IBM_reference, Y2_hexa and AVBP_LW in Table 9), see Figure 13. Note that to compute the MME statistics, a Hodrick-Prescott filter is used to smooth out short-term tracer concentration fluctuations and to compute the MME error and the MME variability based on the trends of the simulated time series at each probe. Note also that the MME error and variability are computed for different sets of probes to analyze their spatial variability: we consider first the whole set of 25 probes to compute the MME error and variability; we then compute the MME error and MME variability only for local probes to focus on near-source dispersion and only for background probes to focus on lateral dispersion (Figure 9). Results confirm that LES models tend to overestimate observations. When removing the first seconds of the time series, the MME variability is more important when looking at best configurations than reference configurations, especially at local probes. This is consistent with the fact that AVBP_LW underestimates the observations in contrary to most LES models of the tested pool, which means that the lower boundary of the envelope significantly changes when moving from the reference configurations to the best configurations.

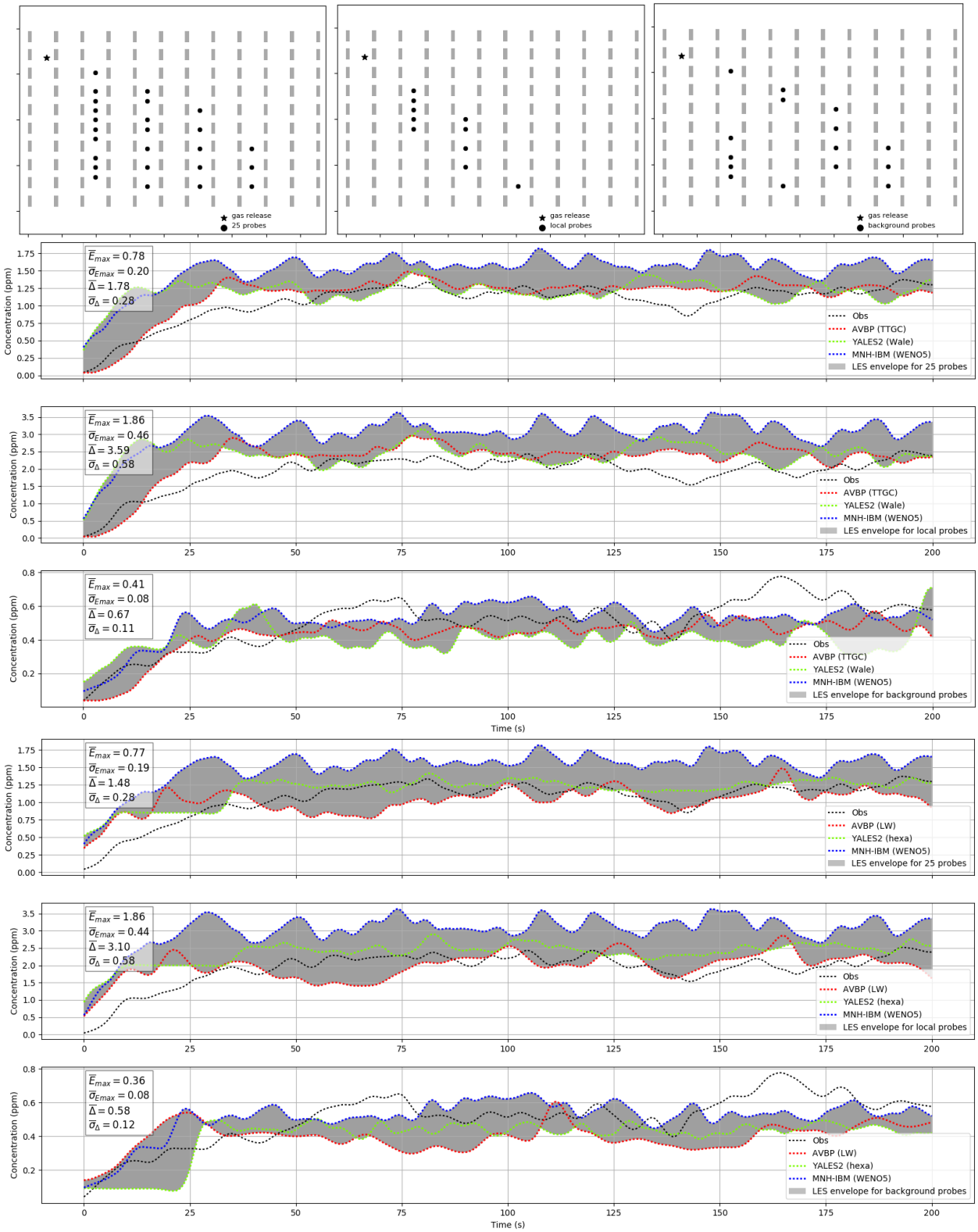


Figure 12: *MME envelopes as a function of time considering different sets of probes for reference configurations (first three panels) and for best configurations (last three panels)*

5) TOWARDS ALEATORY UNCERTAINTY ESTIMATION

Previous results show that the computational grid is one important component when estimating model structural uncertainty. Another important aspect in the LES models, not yet investigated here, is the atmospheric forcing from which the initial and lateral boundary conditions are imposed within the micro-scale computational domain. This can be considered as aleatory uncertainty due to the intrinsic variability of atmospheric flows at meso-scale. Quantifying the impact of this variability on the LES model response was beyond the scope of the present work and will be explored in future work. Still, to prepare this future study and to measure the possible spread in the LES results due to changes in the inlet flow conditions, we carried out four additional simulations per LES code in reference mode with perturbed inlet flow conditions.

In previous simulations, the inlet wind conditions were the same for the pool of eight LES models, with a 4-m wind speed equal to $u_4 = 7.93$ m s^{-1} and a wind direction equal to -41° (Table 1). This corresponds to a friction velocity $u_* = 0.715$ m s^{-1}. The four new configurations correspond to “extreme" values of inlet wind speed and direction; they are set to the mean plus or minus standard deviation extracted from the observed time series in Figure 2. We still use the logarithmic profile (Equation 1) but with a modified friction velocity u_*. To have a 4-m wind speed varying between 6.6 and 10 m s^{-1} as in Figure 2), the friction velocity shall vary between 0.567 and 0.878 m s^{-1}. The corresponding envelope for the logarithmic profiles is presented in Figure 13 (blue dashed lines); the reference logarithmic profile (black solid line) is also represented. The inlet wind direction varies between -29.86° and -47.55° to match wind direction statistics in Figure 2. The envelope for the inlet friction velocity and wind direction are presented in Figure 13. Table 13 summarizes the values of the inputs related to the inlet wind parameterization.

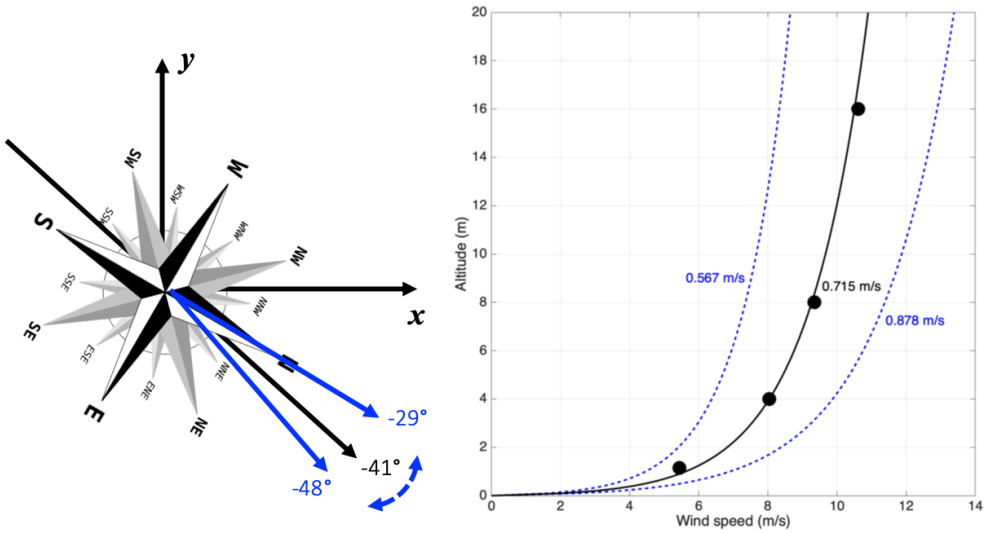


Figure 13: *Envelopes of inlet wind direction and speed based on the flow statistics presented in Figure 2: blue lines correspond to the lower and upper boundaries for the inlet wind direction (right panel) and for the inlet friction velocity (right panel); black lines correspond to the reference configuration.*

Table 13: *Input parameters associated with inlet wind conditions for the “extreme” LES models: α_4 and u_4 are the mean wind direction and the mean horizontal wind speed at 4-m high; u_* is the corresponding friction velocity. These “extreme" values are run for MNH-IBM, YALES2 and AVBP.*

Configuration	u_*	u_4	α_4
Reference	0.715 m s^{-1}	7.93 m s^{-1}	-41°
Extreme 1	0.567 m s^{-1}	6.6 m s^{-1}	-47.55°
Extreme 2	0.567 m s^{-1}	6.6 m s^{-1}	-29.86°
Extreme 3	0.878 m s^{-1}	10.1 m s^{-1}	-47.55°
Extreme 4	0.878 m s^{-1}	10.1 m s^{-1}	-29.86°

Figure 14 presents the potential dispersion (shaded areas) associated with the four extreme simulations in terms of tracer concentration time series at probes 16 and 5. Results show very different sensitivity to the

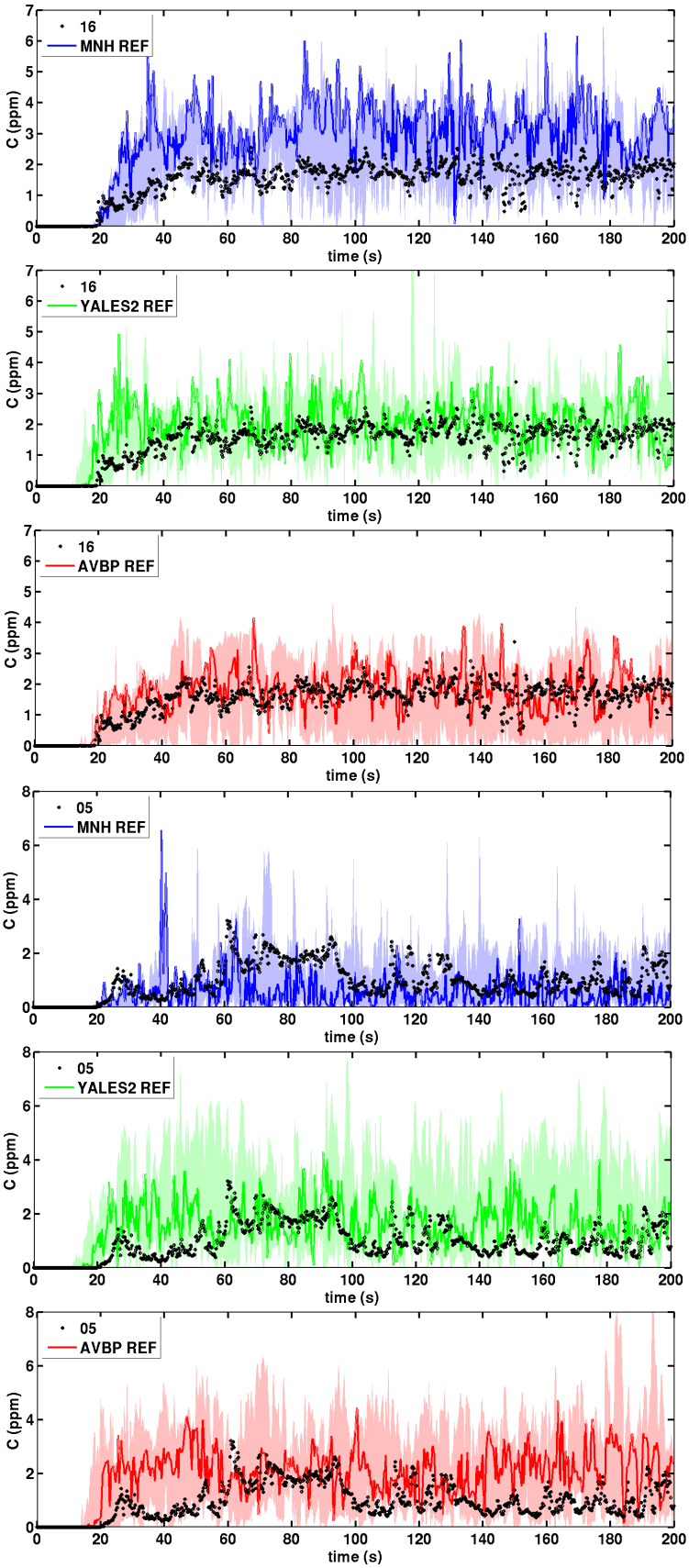


Figure 14: *Time series of tracer concentrations (ppm) at two different sensors (sensor 16 and sensor 5, Figure 1). Measurements are represented using black symbols; reference LES results (MNH-IBM, YALES2, AVBP) are represented using bold solid lines and the perturbed LES results are represented using shaded areas.*

changes of the inlet wind conditions between the three LES codes. For instance, the resulting mean local wind direction changes from 0.2 to 25° in MNH-IBM, from 0.01 to 14° in YALES2, and from 0.6 to 19° in AVBP.

MNH-IBM shows the highest variability among the three codes, with local differences up to 200% at sensors located in the plume (for instance, at sensor 16). Note that MNH-IBM reference simulation does not correspond to the mean value of the envelope. Note also that the wind deviation with respect to the inlet wind direction induced by the containers is more pronounced for the two cases associated with $\alpha_4 = -47.55°$. This implies that starting from a given inlet wind direction, the mean tracer plume is deflected as in the AVBP and YALES2 reference simulations and the channeling effect is no longer as strong as in the MNH-IBM reference simulation. The inlet wind directions in the reference configurations may be near a bifurcation, i.e. a change of plume dispersion behavior for the code MNH-IBM. The presence of such a bifurcation may explain some of the discrepancies between MNH-IBM and AVBP/YALES2 in the model-to-model comparison.

This highlights the importance of measuring the impact of both epistemic and aleatory uncertainties on plume dispersion LES results. Aleatory uncertainties may be even stronger than the aforementioned epistemic uncertainties. Advanced uncertainty quantification methods based on surrogate approaches could be valuable to analyse the impact of aleatory uncertainty in a robust and systematic way and thereby assess the influence of large-scale atmospheric conditions on micro-scale air quality problems. This will be investigated in future work.

CONCLUSION

In the present note, we reported some useful tests to ensure a robust model-to-model comparison for LES of micro-scale meteorology and air pollutant dispersion. A pool of different LES models was simulated for the near-neutral MUST trial 2681829 with three different codes: Meso-NH (anelastic formulation, structured grid, IBM), YALES2 (low-Mach formulation, unstructured grid, body-fitted) and AVBP (compressible formulation, unstructured grid, body-fitted). These codes were used with recommended numerical schemes and models, and with highly resolved computational grids (20-30 cm within the array of containers). In the present note, more information on the LES settings are given, especially for AVBP in order to validate the artificial compressibility approach and to choose outlet boundary conditions that are appropriate to handle compressibility artefacts.

The dependence of the LES results to user's modelling and numerical choices (i.e. grid resolution and type of elements; advection scheme; subgrid turbulence model) was studied to provide a detailed model-to-model comparison and estimate multi-model variability. Various numerical schemes for advection were tested (WENO5, CEN4, TFV4A, TTGC and LW), CEN4 and LW schemes having the property of being more diffusive than their counterparts WENO5 and TTGC. Various subgrid-scale turbulence models were also tested (TKE 1.5, WALE and Smagorinsky). Different computational grids (structured, unstructured, type of cells, spatial resolution) were also tested. In the present note, LES results are reported with a more exhaustive list of plots and statistical metrics than in Rochoux et al. [33] for completeness. Results show that the LES approach is able to represent the short-term plume dynamics. To further analyse the results, statistical metrics were computed at local and background sensors, i.e. at sensors where the tracer concentration is high (above 1 ppm) or low (less than 1 ppm). The LES approach achieves good FAC2 in near-source areas where high tracer concentrations are observed. This indicates a good performance of the LES models evaluated independently. Still, these near-source are also the areas where the multi-model variability is the largest. There are not many measurements in the vicinity of the emission source and in the first layers of the atmosphere. There is thus no way to discriminate the performance of the different LES models in a robust way. This is a limitation of the MUST field campaign. Using high-resolution observation platforms such as unmanned aerial systems or scanning lidar in future field campaigns could be helpful to provide more informative data for model validation.

A preliminary sensitivity study of the LES results to the inlet wind conditions was also carried out to provide further insights into the multi-model variability. MNH-IBM was found to be particularly sensitive to the inlet wind conditions. YALES2 was found to be the least sensitive to the inlet wind conditions but also to model configuration choices.

ACKNOWLEDGEMENT

The financial support provided by the FCS-STAE foundation is greatly appreciated. This work was granted access to the HPC resources from GENCI-TGCC/CINES (2017/A0010110079, 2019/A0062A10822). The authors acknowledge Géraldine Rea and Franck Auguste (previous postdoctorates at CERFACS) for their work to initiate the model-to-model comparison. The authors also acknowledge Bertrand Carissimo (EDF R&D), Eric Pardyjak (University of Utah) and Daniel Cariolle (Météo-France) for helpful discussions and for providing data on both environmental CFD and MUST datasets. They finally acknowledge Benjamin Farcy, Vincent Moureau and Ghislain Lartigue (CORIA) for support on YALES2, Juan Escobar (LA) for support on Meso-NH, Rémy Fransen (CERFACS) for support on Blender visualisation as well as the Computer Support Group at CERFACS.

References

[1] F. Auguste et al. “Implementation of an immersed boundary method in the Meso-NH v5.2 model: Applications to an idealized urban environment”. In: *Geoscientific Model Development* 12.6 (2019), pp. 2607–2633. DOI: 10.5194/gmd-12-2607-2019.

[2] F. Auguste et al. “Large-eddy simulations with an immersed boundary method: Pollutant dispersion over urban terrain”. In: *Atmosphere* 11.1 (2020). DOI: 10.3390/atmos11010113.

[3] C. Biltoft. *Customer report for Mock Urban Setting Test*. Technical Report WDTC–FR–01–121. West Desert Test Center, Dugway, Utah, USA: U.S. Army Dugway Proving Ground.

[4] B. Blocken and C. Gualtieri. “Ten iterative steps for model development and evaluation applied to computational fluid dynamics for environmental fluid mechanics”. In: *Environmental Modelling & Software* 33 (2012), pp. 1–22. DOI: 10.1016/j.envsoft.2012.02.001.

[5] F. Camelli, R. Lohner, and S. Hanna. “VLES study of MUST experiment”. In: *43rd AIAA Aerospace Sciences Meeting and Exhibit*. 2005. DOI: 10.2514/6.2005-1279.

[6] J.C. Chang and S.R. Hanna. “Air quality model performance evaluation”. In: *Meteorology and Atmospheric Physics* 87.1 (2004), pp. 167–196. DOI: 10.1007/s00703-003-0070-7.

[7] P. Colella and P.R. Woodward. “The Piecewise Parabolic Method (PPM) for gas-dynamical simulations”. In: *Journal of Computational Physics* 54.1 (1984), pp. 174–201. DOI: 10.1016/0021-9991(84)90143-8.

[8] O. Colin and M. Rudgyard. “Development of high-order Taylor–Galerkin schemes for LES”. In: *Journal of Computational Physics* 162.2 (2000), pp. 338–371. DOI: 10.1006/jcph.2000.6538.

[9] COST Action 732. *Quality assurance and improvement of micro-scale meteorological models*. COST Office, Brussels, 2009.

[10] F. Couvreux et al. “Intercomparison of Large-Eddy Simulations of the Antarctic Boundary Layer for Very Stable Stratification”. In: *Boundary-Layer Meteorol.* 176.3 (2020), pp. 369–400. DOI: 10.1007/s10546-020-00539-4.

[11] J. Cuxart, P. Bougeault, and J.-L. Redelsperger. “A turbulence scheme allowing for mesoscale and large-eddy simulations”. In: *Quarterly Journal of the Royal Meteorological Society* 126.562 (2000), pp. 1–30. DOI: 10.1002/qj.49712656202.

[12] A. Dejoan et al. “Comparison between large-eddy simulation and Reynolds-averaged Navier–Stokes computations for the MUST field experiment. Part II: Effects of incident wind angle deviation on the mean flow and plume dispersion”. In: *Boundary-layer meteorology* 135.1 (2010), pp. 133–150. DOI: 10.1007/s10546-010-9467-2.

[13] J. Franke et al. “The COST 732 best practice guideline for CFD simulation of flows in the urban environment: A summary”. In: *International Journal of Environment and Pollution* 44.1-4 (2011), pp. 419–427. DOI: 10.1504/IJEP.2011.038443.

[14] N. Gourdain et al. “High performance parallel computing of flows in complex geometries: I. Methods”. In: *Computational Science & Discovery* 2.1 (2009), p. 015003. DOI: 10.1088/1749-4699/2/1/015003.

[15] V. Granet et al. “Comparison of Nonreflecting Outlet Boundary Conditions for Compressible Solvers on Unstructured Grids”. In: *AIAA journal* 48.10 (2010), pp. 2348–2364. DOI: 10.2514/1.43729.

[16] S.R. Hanna, O.R. Hansen, and S. Dharmavaram. “FLACS CFD air quality model performance evaluation with Kit Fox, MUST, Prairie Grass, and EMU observations”. In: *Atmospheric Environment* 38.28 (2004), pp. 4675–4687. DOI: 10.1016/j.atmosenv.2004.05.041.

[17] M. Kraushaar. “Application of the compressible and low-Mach number approaches to large-eddy simulation of turbulent flows in aero-engines”. PhD thesis. Institut National Polytechnique de Toulouse (INPT), 2011.

[18] C. Lac et al. “Overview of the Meso-NH model version 5.4 and its applications”. In: *Geoscientific Model Development* 11.5 (2018), pp. 1929–1969. DOI: 10.5194/gmd-11-1929-2018.

[19] J.P. Lafore et al. “The Meso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations”. In: *Annales Geophysicae*. Vol. 16. 1998, pp. 90–109. DOI: 10.1007/s00585-997-0090-6.

[20] M. Lesieur and O. Metais. “New trends in large-eddy simulations of turbulence”. In: *Annual Review of Fluid Mechanics* 28.1 (1996), pp. 45–82. DOI: 10.1146/annurev.fl.28.010196.000401.

[21] X.-D. Liu, S. Osher, and T. Chan. “Weighted Essentially Non-oscillatory Schemes”. In: *J. Comput. Phys.* 115.1 (1994), pp. 200–212. DOI: 10.1006/jcph.1994.1187.

[22] T. Lunet et al. “Combination of WENO and explicit Runge-Kutta methods for wind transport in Meso-NH model”. In: *Monthly Weather Review* 145.9 (2017), pp. 3817–3838. DOI: 10.1175/MWR-D-16-0343.1.

[23] M. Malandain, N. Maheu, and V. Moureau. “Optimization of the deflated Conjugate Gradient algorithm for the solving of elliptic equations on massively parallel machines”. In: *Journal of Computational Physics* 238 (2013), pp. 32–47. DOI: 10.1016/j.jcp.2012.11.046.

[24] V. Masson et al. “The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes”. In: *Geoscientific Model Development* 6 (2013), pp. 929–960. DOI: 10.5194/gmd-6-929-2013.

[25] J. Meyers. *Quality and Reliability of Large-Eddy Simulations*. Ed. by Bernardus J. Geurts. ERCOFTAC series ; 12. First QLES meeting on Quality and Reliability of Large-Eddy Simulation, held October 22–24, 2007 in Leuven (QLES07). Springer, 2008. ISBN: 9781402085772. DOI: 10.1007/978-1-4020-8578-9.

[26] M. Milliez and B. Carissimo. “Numerical simulations of pollutant dispersion in an idealized urban area, for different meteorological conditions”. In: *Boundary-Layer Meteorology* 122.2 (2007), pp. 321–342. DOI: 10.1007/s10546-006-9110-4.

[27] V. Moureau, P. Domingo, and L. Vervisch. “Design of a massively parallel CFD code for complex geometries”. In: *Comptes Rendus Mécanique* 339.2-3 (2011), pp. 141–148. DOI: 10.1016/j.crme.2010.12.001.

[28] F. Nicoud and F. Ducros. “Subgrid-scale stress modelling based on the square of the velocity gradient tensor”. In: *Flow, turbulence and Combustion* 62.3 (1999), pp. 183–200. DOI: 10.1023/A:1009995426001.

[29] T.J. Poinsot and S.K. Lele. “Boundary conditions for direct simulations of compressible viscous flows”. In: *Journal of Computational Physics* 101.1 (1992), pp. 104–129. DOI: 10.1016/0021-9991(92)90046-2.

[30] Y. Qian et al. “Parametric sensitivity and uncertainty quantification in the version 1 of E3SM atmosphere model based on short perturbed parameter ensemble simulations”. In: *Journal of Geophysical Research: Atmospheres* 123.23 (2018), pp. 13, 046–13, 073. DOI: 10.1029/2018JD028927.

[31] L Quartapelle and V Selmin. “High-order Taylor-Galerkin methods for nonlinear multidimensional problems”. In: *Finite Element Fluids* 1375 (1993).

[32] J.D. Ramshaw, P.J. O’Rourke, and A.A. Amsden. *Acoustic damping for explicit calculations of fluid flow at low Mach number*. Technical Report LA–10641-MS. Los Alamos, New Mexico, USA: Los Alamos National Laboratories, 1986.

[33] M.C. Rochoux et al. “Large-eddy simulations of micro-scale pollutant dispersion: multi-model variability for the MUST experiment”. Submitted to Boundary-Layer Meteorology.

[34] J. L. Santiago et al. “Comparison between large-eddy simulation and Reynolds-averaged Navier-Stokes computations for the MUST field experiment. Part I: Study of the flow for an incident wind directed perpendicularly to the front array of containers”. In: *Boundary-Layer Meteorology* 135.1 (2010), pp. 109–132. DOI: 10.1007/s10546-010-9466-3.

[35] P. Schmitt et al. “Large-eddy simulation and experimental study of heat transfer, nitric oxide emissions and combustion instability in a swirled turbulent high-pressure burner”. In: *Journal of Fluid Mechanics* 570 (2007), pp. 17–46. DOI: 10.1017/S0022112006003156.

[36] T. Schönfeld and M. Rudgyard. “Steady and unsteady flow simulations using the hybrid flow solver AVBP”. In: *AIAA journal* 37.11 (1999), pp. 1378–1385.

[37] J. Smagorinsky. “General circulation experiments with the primitive equations: I. The basic experiment”. In: *Monthly Weather Review* 91.3 (1963), pp. 99–164. DOI: 10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2.

[38] E. Yee and C.A. Biltoft. “Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles”. In: *Boundary-Layer Meteorology* 111.3 (2004), pp. 363–415. DOI: 10.1023/B:BOUN.0000016496.83909.ee.