This ms examines coupled aerosol-chemistry-dynamics within a street canyon using a chemistry-aerosol box model, SSH-Aerosol, and two CFD models, OpenFOAM and Code_Saturne. The model configurations are intended to mimic conditions within a specific street canyon in Paris during a field campaign in 2014. It is shown that the coupled model leads to significant differences compared to a baseline simulation in which chemical and physical transformations are excluded.

Given that the vast majority of the urban pollutant dispersion literature assumes passive scalar dynamics, there is certainly a need for more studies of coupling, especially in the case of aerosols, which have received much less attention than gas-phase photochemistry. Nevertheless, several such studies have already appeared, e.g. Sanchez et al. (2016) [1], Han et al. 2018 [2], Gao et al. 2022 [3]. The authors should therefore highlight the novel aspects of their study. They should also discuss the physical or chemical mechanisms underlying their findings. While many sensitivity tests and models comparisons are presented, I did not get the impression that the ms attempts to address a basic physical question.

I also have several technical concerns/questions.

Major points

1. The CFD models are rather idealised.
    (a) I have doubts about the restriction to a 2-D street canyon. While it's well-established that flow within a street canyon is approximately two-dimensional, at least when the external flow is perpendicular to the canyon axis, this does not necessarily mean that a 2-D model works equally well for coupled simulations. Chemical and physical transformations depend on the residence time, which may differ for a fully three-dimensional flow. My hunch is that the residence time for a 3-D canyon would be shorter. Thus the effects of chemical reaction or aerosol processes could be weaker than what the authors report.
    (b) A RANS closure may not be sufficiently accurate. While RANS models agree well with large-eddy simulation when it comes to mean statistics, discrepancies are typically larger for second-order quantities. I suspect that there is a similar phenomenon for pollutants governed by nonlinear evolution equations.

    The authors should discuss the limitations of their models more carefully. How would their major findings be affected if a more realistic or accurate model (e.g. 3-D LES) were used?

2. The CFD models are not validated. This means that the results of, e.g., the model evaluation (Sec. 4), need to be interpreted carefully. On account of nonlinearity, the differences between CFD-chemistry and CFD-passive may not be due only to chemistry or aerosol processes. For example, the limitations discussed in Point 1 could also play an important role.

3. I don't understand how the time-dependent inflow is implemented.
    (a) It's mentioned that flow is driven by linearly interpolated time series of friction velocity and temperature (Figure 2), but how exactly is this done? Is there a time-dependent boundary condition? If so, how often is it updated? Are the authors sure that this procedure is robust?
    (b) Is the temperature a prognostic variable? It's unclear whether the simulations are neutral or not. An inlet velocity profile is specified (eq. 1) but the temperature profile is unspecified. If neutral stability is assumed, what's the justification for this?

(c) The specification of the velocity profiles ignores the dependence on wind direction. The authors assume that the external wind is always perpendicular to the canyon axis. For an application to a real street canyon, this could be a major omission.

(d) The description of the operator splitting seems incorrect. The authors claim that the symmetric Strang scheme [4], $CFD(\Delta t/2) - Chemistry(\Delta t) - CFD(\Delta t/2)$ is second order while the simple alternation of CFD and chemistry, $CFD(\Delta t) - Chemistry(\Delta t)$, is first order. There are a few problems with this.

   i. Strang's result actually requires that the numerical schemes for the decomposed substeps be second order (see p. 510 of his paper). The point isn't that the composition increases the order but rather maintains it. This is reasonable. Strang's method can be applied to general PDEs of the form $u_t = a + b$, where a and b are functions. If a and b are independent of each other but discretised using a first-order scheme, it seems rather unlikely that the composition should be of higher order.

   ii. Strang's result is for a PDE that's been decomposed into functions of independent variables, e.g. $u_t = a(Du, x, t) + b(Du, y, t)$, where x and y are the usual space variables and Du represents spatial derivatives. The authors have equated the terms a and b with the chemistry and CFD substeps. I don't think this is the same thing. They're applying the method to a PDE of the form $u_t = a(Du_{CFD}, x, y, t) + b(u_{chem}, x, y, t)$ where the *dependent variables* u_{CFD} and u_{chem} denote subsets of the original solution space (i.e. $u = u_{CFD} + u_{chem}$). The operator splitting considered by the authors is conventional in the numerical solution of turbulent reacting flows, but it's different in kind from the splitting analysed by Strang (which generalises alternating direction methods).

   iii. I think it's misleading to imply that a higher-order scheme necessarily yields a more accurate solution of a turbulent reacting flow. The splitting should depend on the coupling between dynamics and chemistry, i.e. the Damkohler number. Thus a method, like the authors', which has a chemistry sub-timestep that's longer than the dynamical sub-timestep, will not be accurate in certain limits (e.g. slow chemistry) regardless of what the overall order might be.

   iv. If the Strang scheme is in fact higher-order or more accurate than the simple first-order scheme, then why should the simulated concentrations be largely insensitive to the timestep (Fig. 10)? This would imply that the actual errors are not described by classical error analysis, which is based on linearisation. If so, there's little point in referring to the order of the schemes.

4. The results are limited to a 12-hour simulation for a single day. Why? The authors should show that their results are robust by considering different days.

Minor points

1. The English should be checked carefully. There are many small errors. I suggest that the singular be used in place of 'inorganic and organic matters', etc.

2. The organisation of the ms could be improved. Here are several examples:

   (a) The Abstract is too long and jumps between paragraphs.

   (b) In the Introduction, the role of vehicles in emitting ammonia is mentioned after the OpenFOAM and Code_Saturne are introduced.

    (c) Grid details aren't provided with the model description (Sec. 2) but deferred to Sec. 4.4

3. l.121 Why are two CFD models considered? One would expect well-tested models to give similar results. Why is the model comparison of general scientific interest?
4. l.133 Why is deposition implemented in the CFD models rather than SSH-aerosol? Since the chemistry and dynamics timesteps differ, this could introduce an inconsistency between deposition and other aerosol processes.
5. l.143 Why isn't nucleation considered?
6. l.144 Why were these bound diameters chosen?
7. l.173 Fig. 2 I presume that these figures correspond to a specific day. Why was this day chosen? Measurements were conducted from 6 April to 15 June.
8. l.191 ff. How is the emission spectrum of the aerosols defined?
9. l.204 What is the justification for setting the pressure to zero at the outlet? This choice could affect the flow within the canyon if the domain isn't sufficiently large. The authors should probably test the sensitivity to the domain size and/or perform a validation against wind-tunnel data.
10. l.263 Table 1 Some of the errors for CFD-chemistry are rather small. In light of the idealised numerical configuration, this seems fortuitous. It appears that the statistics correspond to a canyon average (i.e. Fig. 4). However, measurements are limited to a few points. Is this a fair comparison?
11. l.306 "This method is called the constant-condition method (CCM) in this study, in contrast to the transient-condition method (TCM)." As mentioned above, how is the transient-condition method implemented? Note that this information should be included in the model description.
12. l.318 "However, it should be noted that CCM cannot replace TCM when simulating long periods because the mass concentration may not change linearly between the selected time points." I don't understand this.
    (a) I think the issue isn't the length of the simulations but rather the representativeness of the inlet boundary conditions vis-a-vis the actual inflow.
    (b) Why is the mass concentration the only issue? In practice, variations in the inflow wind speed and direction play a crucial role.

References

1. Sanchez, B., J.-L. Santiago, A. Martilli, M. Palacios, and F. Kirchner, 2016: CFD modeling of reactive pollutant dispersion in simplified urban configurations with different chemical mechanisms. *Atmospheric Chemistry and Physics*, 16, 12143–12157.
2. Han, B.-S., J.-J. Baik, K.-H. Kwak, and S.-B. Park, 2018: Large-eddy simulation of reactive pollutant exchange at the top of a street canyon. *Atmospheric Environment*, 187, 381–389.
3. Gao, S., M. Kurppa, C. K. Chan, and K. Ngan, 2022: Technical note: Dispersion of cooking-generated aerosols from an urban street canyon. *Atmospheric Chemistry and Physics*, 22, 2703–2726.

4. Strang, G., 1968: On the Construction and Comparison of Difference Schemes, *SIAM J. Numer. Anal.*, 5(3), 506–517.