General comments:

The study "**A Regional multi-Air Pollutant Assimilation System (RAPAS v1.0) for emission estimates: system development and application** by Shuzhuang Feng and colleagues describes a proposed method to estimate chemical and particulate matter emissions from observations by inversion. Emission rates of trace gases both of biogenic sources and anthropogenic chemical pollutants cannot be measured directly (apart from very special cases). Yet by evident reasons a precise knowledge is utterly important. While it is a long-discussed topic in atmospheric chemistry data assimilation and inversion, a solution of which is hampered by other factors, as initial and boundary data, deposition rates and many other parameters exert significant influence on the simulation result of chemistry transport models (CTM), which are the link between observations and emission estimates.

As their central objective the authors claim this study offer "a useful tool for accurately quantifying multi-species anthropogenic emissions at large scales and near-real time, which will serve better for monitoring emission changes and designing future emissions regulations and pollution control."

While the aspirations of the authors are remarkable, I could not find sound justifications for both a validated methodological approach and the practical proof of concept by the presented case study. I cannot recommend publication prior to a presentation of supporting evidence of the central claim.

Specific comments:

1. The introduction comprises related literature on emission inversion, however with a clear bias toward Kalman filtering. The variational approach, which plays a key role in radiatively active trace gases for greenhouse effects, is also used in reactive chemistry inversion. Yet this methodological branch is disposed of by remarking that this is "technically difficult and cumbersome for complex chemical transport models", without addressing its features. Literature survey may be complemented to a appropriate level here.

2. A central position next to emission optimization is initial value optimization as a prerequisite for unbiased emission rate estimates. A discussion on sensitivities of CTMs to additional parameter controls (see above), which demonstrate the authors' awareness of competing impacts on model simulations is lacking however. As it is presented, it is tacitly assumed that these sensitivities are minor. In this context, the authors' approach may well be valid. But it is a matter of scientific scrutiny to meticulously expose the underlying assumptions, which would not affect the value of the study.

3. Methodology

3.a Inversion and data assimilation methods perceived in the referenced literature address Best Linear Unbiased Estimators (BLUE). This property should also be granted by the two-step procedure combining three-dimensional variational (3DVAR) algorithm and the ensemble square root filter (EnSRF) algorithm, which is claimed to show "that the "two-step" scheme clearly outperformed the simultaneous assimilation of ICs and emissions ("one-step" scheme), …" (line 72). While splitting tasks into sequential steps is a well established strategy in numerics to reduce complexity and increase efficiency (e.g. ADI methods), essential care must be taken to ensure the convergence of both split (two-step) and combined (one-step) algorithms toward the same result. The claimed

robustness of the method is demonstrated here not in a sense of same result. Rather different background emissions are probed for convergence (e.g. line 153), which does not experimentally validate the two-step method. Moreover, it is a problem, that the two-step method is established by two different methods. Unless there is a sophisticated multivariate (that is multi-species cross-correlating) background error covariance matrix (which is evidently not the case in this study) initial value estimation by 3DVAR cannot account for some realistic chemical consistency. Even monovariate cross-correlations are essential to be defined, both horizontally and vertically, to optimize unobserved adjacent areas/height levels. Any estimation errors made in step one are compensated by step two provided biased emission optimization, which is, by EnKF, also performed in time. (Say, too high/low estimated vertical concentrations aloft and mixed down to the observation site are adjusted by too low/high emissions). So the system is prone to propagate errors made in a single step, the methodologically weakest of which limits the overall success. So better analyse why (see L 71-72) it can be possible that "the "two-step" scheme clearly outperformed the simultaneous assimilation of ICs and emissions ("one-step" scheme)".

The authors should either proof the BLUE property of their two-step method with a unique solution by mathematical rigour, or adopt a one-step procedure based on their EnKF approach, or even better, an ensemble Kalman Smoother.

3.b In addition, it is a well established validation procedure and common practice after implementation of a new data assimilation/inversion algorithm, to test the system by "identical twin" experiments and Observation System Experiments (OSEs), where a virtual reality is given by "nature runs", where "artificial observations" serve to estimate known chemical states and emission rates (see textbook by Daley, 1990 for details). There is no hint in the paper, that corresponding activities have been undertaken. These tests do not confirm the correctness of the theoretical approach. This being assumed, they seek a necessary, yet insufficient test being approved. The authors are strongly encouraged to validate and proof their approach by these tests.

4. The presentation of the EnKF is confusing (see equs. (7)-(12)): Below (7) we read "δX^b_i represents the randomly perturbed samples that are added to the prior emissions X^b_0 to produce ensemble samples of the inputs X^b_i …..". Evidently these parameters are <u>emissions</u>. In equs (8) and (11), also others in between, these parameters are used as state parameters in traditional ensemble KF notation, revealing in (8) a comparison with observations ($y-HX^b$). I presume observation operator H does not link emissions X with observations y. Rather concentrations are surely meant here (and not the extremely unlikely case of available eddy covariance tower flux measurements). A means to unravel this would be the approach of expanding the state vector of concentrations by emission rates following Wu at al. (2016). See their eq. (5.3). Take also note of their discussion of optimization and control on the finite assimilation window length. What is the assimilation frequency within a DA window of 1 day? Please clarify these points.

5. Is there a test set demonstrating prognostic improvements? And if so, which forecast time is used. I was unable to identify such an experiment, which demonstrates a type of benefits from emission optimization.

6. L 203: The grid size of 36 km is very coarse. In fact to be judged in relation to your following statement:

L 878-883 "In additionally, O3 observations are not assimilated to improve NOx and VOC emissions using cross species information due to the strong nonlinear effects within the O3-NOx-VOC relationship, in which the O3 concentration and NOx (VOC) emissions are positively correlated in the

NOx (VOC)-limited region and negatively correlated in the VOC (NOx)-limited region (Tang et al., 2011)." NOx chemistry is linked to VOCs and CO via O3. Due to the lack of VOC observations, the evolution of O3 (build up) is mandatory, as otherwise NO2 evolution is not properly analysable. This clearly indicates that the model set-up/coarseness is unable to simulate the core reactivity of tropos. chemistry properly and thus is unsuited to infer emission rates. It is suggested to rerun the experiment with asufficiently highly resolved nest on a densely observed area.

Technical corrections:

L. 59-63: What is meant in one case: Emission optimization with or without prior IC by 3D-var?

L. 68-69: Selection of the Dec 2016 case study analysis: in view of typically episodic nature of mineral dust: does this high quantitative increment make sense for a validation of a novel algorithm? Why not another episode?

L 122-124: Unclear: "Barbu et al. (2009) updated sulfur oxide (SOx) emissions with SO2 gas and sulfate aerosol observations and showed that forecasts were improved overall but degraded when derived only from SO2 or sulfate observations." Better formulate logics: ... do you mean alone?

L 176-177: "It runs only once and provides a "perfect" chemical ICs for the subsequent EI subsystem." Justify "perfect". Justify analysis of unobserved height levels?

L. 180-181: "which are then sampled according to the locations and times of the observations" Unclear, more technical details are needed here.

L. 184: Do you mean DA window length of 1 day?

Figur1 a caption too short. More detail on precise times and data update frequency in the En KF needed.

L 225 "represent the measurement sites" Do you mean "chemistry"?

L 242: "… of surface air pollutant observation operators,…" Explain please.

L 312 "the horizontal length scales decrease with increasing heights," Please explain why, and not the opposite.

L 314 " The ground-level scale generally spread 40-45 km for all control variables on average." This is little more than a grid cell. How defined?

L 315-318: More detailed explanation required: Doo you mean stack overshooting vs. surface emissions? A result of NMC? Is boundary height and related mixing considered?

L 323-324: Whitaker and Hamil is a meteorological application without emission optimisation. For square root filtering technique fundamentals please refer to much earlier textbook literature, eg. Bierman: 1977, or Maybeck 1979.

L 336 variable localization: meaning?

L 336-337: Do you mean also at the same location?

Eq (12) Matrix in a denominator? Please explain.

L. 384-386: This is too much a try-and-error like discussion. Emissions are not constant or diurnially constant over these times: weather changes, working days, holidays, ... Please be more precise about your reasoning.

L425 arithmetically

L 446 are distributed

L 449 unrepresentative

L 452-453 Explain relations, formulae better.

L 461 25/336 is a fairly low fraction. Please justify.

L 534- 540 For emission inversion the boundary layer height and its prevalent stability is more decisive than any other parameter. Can you please make efforts to validate this by radiosonde data, if available?

L 559 "These statistics indicate that the initial fields can be adjusted effectively by our IA subsystem." I think this claim is not sustained, despite the fact that emissions are not yet corrected: What about height levels above? How are they validated? How can be made sure that CB5 chemsitry fields are in the proper chemical balance of the system (notably NO_x, O3, CO, VOCs)?

L 619: This applies more to the NOx-O3-CO/VOC chemistry rather than PM2.5 .

L 662: Where did you get the \sigma from? For KF an analysis error covariance matrix A should be available for this.

L 701-702: With 1000% excess is linearisation still acceptable? A scrutinized analysis is appropriate.

L 705 "In addition, without dust may be another reason, since no wind blowing dust scheme was applied in this study as mentioned above." Check English meaning.

L727-728 : Please demonstrate this statement.

L 746-747: This is well known, but demonstrate that this flaw cannot happen. Please my mathematical means.

Fig 13 emission changes

Literature:

Bierman: Factorization methods for Discrete Sequential estimation, Academic Press, 1977, or

Maybeck: Stochastic Models, Estimation and Control Academic Press, 1979.

Wu, Xueran, B. Jacob, and H. Elbern: Optimal control and observation locations for time-varying systems on a finite-time horizon, SIAM J. Control Optim., 54, 291-316 , 2016.