General comments:

In this work a hybrid model has been developed and evaluated to analyse the effects of vehicle emissions on urban roadside concentrations of NO2 in Beijing. The article is well written and raises an important topic, the link between the simulations done using regional chemistry transport models and the simulations at urban level done using gaussian/dispersion models. However, there are few points that should be clarified in order to make clearer the evaluation of the model and the scenarios tested.

Major comments:

- The introduction clearly shows the differences between chemistry transport models and dispersion/gaussian models highlighting the difficulties of the former in predicting the roadside concentrations. However, there isn't a clear link between regional models and urban models. Few works have been published and few models have been already developed to couple regional and urban models and these should be mentioned in the introduction
- The methodology highlights only part of the process defined in Figure 1. The authors focus their discussion on the urban model but WRF and CMAQ configuration and outputs should also mentioned and discussed.
- The simulations are run for a period of high photochemical activity. This is surely dependent by weather conditions that are completely absent from the article.
- The NOx-O3 system include also VOCs. There is no mention of this in the methodology or in the results. Are VOCs included in the simulations? I t would be good to add the chemical mechanism somewhere in the supplementary material.

Minor comments:

Line 32: the reference (Cui et al., 2021), (Shah et al., 2020) should be (Cui et al., 2021; Shah et al., 2020)

Line 33: delete that: in is still much more severe than ~~that~~ in developed

Line 40 – 42: The comparison with the emission in Lyon is quite specific. I suggest to explain a bit more or in alternative to make a more general case of "other urban areas".

Line 102: the spatial resolution should be precise: please substitute < 100 m x 100 m with the real spatial resolution.

Line 107 -108: The choice of the midpoint height 22.5m suggests that the CMAQ has a first vertical layer at 45m of height. In first instance this would be in my opinion too high. Generally, CTMs has the first 9-10 vertical layers below the boundary layer but, to improve the prediction on the ground level, keep the 1st layer around 20m from the ground. Could the authors enforce their statements with one or more references that justify this choice?

Line 277 – 279: The authors describe the performance of the model in terms of "high" and "low" RE. It would be good to provide a more quantitative description or a reference value for this particular metric.

Line 280 – 283: I'm not sure that the MARS model performs better than RD in Figure 5. I suggest to clarify this paragraph better. In the a) figure the CFD is the closest to the observations, followed by RF

(red slope) and MARS (yellow slope). In figure b) again CFD is the closest to the observations, RF is completely underestimated and MARS overestimated from values of z/H > 0.25

Line 378 – 383: The authors mentions NOx emissions leading to high NO2 observations among all sites. They also say that CMAQ model underestimate the NO2 concentrations near ring roads (MB = -15µg/m3). The NOx emissions account for NO+NO2, if this variable is NOx before to be inserted in CMAQ, it has to be divided between NO and NO2. In roadside sites generally the NO emissions are high, could the underestimation in CMAQ related a not precise division of the original emissions of NOX in NO and NO2?

Line 390 – 394: The model actually improves the performance of O3 in comparison with CMAQ only model. This agrees with the underestimation in NO2 that the CMAQ only shows and previously described by the authors. Being the O3 chemistry dependent not only by NO, NO2 but also by VOCs I would spend some words to introduce these pollutant class and give more details on them.