The study addresses an interesting issue relevant with the performance of European scale models in the representation of the chemical composition of PM in the near surface air. The scientific tools used to address the issue are well documented and methodology is scientific sound. Clarifications are necessary to describe better the methodological steps. Also, the paper provides some answers on the reasons for the better or worse model performance; however most of them have been discussed in previous publications. Also the uncertainties in the model evaluation presented in the paper are only theoretically addressed but not quantified. Interesting is the issue of the water mass included in the aerosols, which in most cases is not accounted for in the validation of PM model results; this issue is also theoretically discussed in the
Following are some suggestions and comments for the submitted manuscript:

1) Section 2.3: The authors should present a table in which all the emitted and simulated PM species will be presented in detail for each model. The existing Table 2 cannot explain how sea salt is speciated in model simulations (is it simulated explicitly as Na and Cl?). Similar is the comment for fire PM species. Usually, fire PM emissions are simulated as OC and EC emissions and I would expect these emissions to have a contribution to OC and EC levels. How is this issue addressed in the different models? If fire PM emissions were speciated as OC and EC then how was it possible to distinguish between anthropogenic OC/EC and fire-related OC/EC? The models simulate OC or OA? Also in Table 2, the PM size is not presented (PM10 versus PM2.5).

2) Section 2.4: Please present in Table 3 the size of the PM measured (PM10 versus PM2.5).

3) Section 2.5: The method for the estimation of nss-Ca levels from the model results is not clear. Are the simulated nss-Ca concentrations estimated as the sum of the 10% of dust concentrations plus 3.5% of the unspeciated other primary PM concentrations? The authors should also make clear in the whole manuscript (e.g. in Table 6) that it is not the mineral dust model results that are evaluated but the nss-Ca values (including both the desert dust and the anthropogenic contribution).

4) Section 3.1, page 9, lines 7-18: The authors explained the bad model performance in Schauinsland Mountain station as the result of model spatial resolution and high station altitude. The position and altitude of the stations should be provided in the supplement. Are there other stations of similar altitude in which the models performance is not as bad in the Schauinsland Mountain station?

5) Section 3.1, page 10, lines 2-4: Explain the reasons for models’ different ability in simulating PM2.5 and PM10 in summer and winter.
6) Section 3.1, page 10, line 11: The authors state: “The medianComp fully includes SOA, desert dust and fire-induced PM.” This is not true since according to Table 2, SOA are not included in LOTOS-EUROS and in SILAM. Also BCs dust is not included in LOTOS-EUROS.

7) Section 3.1, page 10, lines 16-18: The authors explain that the differences between median and medianComp are due to desert dust. However, according to Table 2, it is only the fire originated PM that were not included in total PM while BCs dust was included all CMAQ, EMEP and SILAM.

8) Section 3.2, Table 6: Which are the size bins of the PM species validated (comment mostly for sea salt and mineral dust). The words “mineral dust” should be replaces with “nss-Ca”. Why evaluation is presented explicitly only for SO2 and not for other PM precursor gas compounds like NH3 or NOx?

9) Section 3.2.3: The performance of CMAQ, can it be explained by the fact that fire emissions are included in other PM?

10) Section 3.2.5: Any comments on OA comparison with observations should be based on CMAQ and EMEP results since the other models include OC in other PM. Similar remark for dust straightforwardly included only in EMEP and SILAM.

11) Section 3.2.5., page 15, line 7: Correct OA with OC.

12) Section 3.2.5., page 15, line 11: Please add that also the OC is for some models included in primary aerosols (as the case for LOTOS-EUROS and SILAM)

13) Section 3.2.5., Figure 8: a) suggestion: add the simulated SOA and POA and compare them with the observed TOA, b) why fire PM are presented separately and not speciated in OC and EC so as to allow better comparison with observations? (see also comment 1). In the legend of Figure 8 the PM species are not presented in the same order as they appear in the plots.

14) Section 4.1: This section provides a theoretical description of the possible rea-
sons for overestimations or underestimations in the model results. However, there is no quantification of the uncertainties. For example, how much is the uncertainty introduced in the model validation because of the assumption that dust and anthropogenic mineral aerosols have a 10% and 3.5 Ca content respectively? Clarifications also are necessary in the last paragraph of this section in which the authors discuss about the absence of fire emissions from the computations while in Table 2 fire originated PM are presented as included in the simulations of all models.

15) Section 4.3: It is a very interesting section. The authors should present results on the models’ improved performance when water is accounted for (even in the four stations providing a complete set of PM measurements; after all section 3.2.5 was based on the measurements in these few stations).

16) Figure 6: Correct in the lower-right plot the OC to EC.

17) Figure S6: Was CMAQ excluded from the average since dust in CMAQ was included in other PM?

18) Figure S7: Was LOTOS-EUROS and SILAM excluded from OA mean values since OA in these modes were included in other PM? The Figure is not commented in the text of the manuscript.

19) Figure S8: Please check comment 1. How fire emissions were simulated and consequently presented in the Figure for PM? If speciated to OC and EC emissions, how fire-related OC and EC concentrations were distinguished from anthropogenic OC and EC concentrations?