Replies to the comments to the manuscript '*Advanced error diagnostics of the CMAQ and Chimere modelling systems within the AQMEII3 model evaluation framework*' by Solazzo et al., 2017(doi:10.5194/acp-2017-257)

Reviewer 2

General Comment: This study pursues to outline the evaluation methods applied during the three phases of the AQMEII activity in view of building an evaluation strategy. Annual simulations of the WRF-CMAQ and IFS-Chimere are evaluated, as well as runs with perturbed emissions or boundary conditions or ozone deposition. The analysis focuses on ozone simulations over selected European and North American subregions. The validation is based on (a) the decomposition of the mean square error into bias, variance and covariance (to isolate the sources of error into systematic, variability and phasing) and (b) the decomposition of the time series into spectral components (to investigate the temporal characteristics of the error and elicit the associated misrepresented processes). The authors conclude that both decompositions can aid understanding of the causes of model error, when sensitivity simulations or process analysis simulations are not available. The study is addressing an interesting problem and I recommend it should be published.
We thank the reviewer for the supportive comments.

Specific Comments

- [Eq 1, L174] Please include some explanations for Eq 1. Is it an identity? Are there any assumptions behind it?
  The derivation of Eq.1 is standard and can be found in textbooks as well scientific publications. In the context of AQMEII, it has been extensively discussed in previous publications. It is an identity, i.e. is not an approximation. For self-consistency, a short description of the terms has been provided in the revised manuscript.
- [ACF & PACF discussion, paragraph 4.2] Please include in the text only arguments arising from significant correlations. In the same manner, significant correlations should be distinguishable in Figures 15 and 16.
  Statistically speaking, the auto and partial correlations reported in Figures 15 and 16 are significant when lying outside the horizontal blues lines centred at zero and having width proportional to the 95% significance level (i.e. threshold for the null hypothesis to be rejected of 0.05): $r_{0.95}=0 \pm 2N^{0.5}$, where N is the sample size. Therefore, the values of the auto-correlation are significant (although small sometimes) throughout the range of lags shown in Figure 15 and 16, while the partial correlation are small and non-statistically significant for lags larger than 60 hours (when the diurnal fluctuations are removed). We could therefore reduce the x-axis of the PACF plots of the 'non-diurnal fluctuations' plots, but for consistency of the discussion and presentation of the results we would rather keep the same format.
- [kz filter, paragraph 4.3] Please provide an estimate for the error leakage from the use of non-independent spectral components and discuss its impact on the results.
  The effect of the leakage among components (and its quantification) for the models participating in AQMEII is extensively discussed in Solazzo and Galmarini (2016), Hogrefe et al. (2013), Galmarini et al. (2013), Kioutsioukis and Galmarini (2016) and Solazzo et al., (2017), but it is a sensitive issue and we have added a discussion in the revised text. In this study we use the kz filtering only for the purpose of isolating the portion of the ozone time series that is faster than ~1.5 days. The effect of leakage is rather limited for this particular application (5 to 10%, Figure 1 of Solazzo and Galmarini 2016).
- [regression, paragraph 4.4] The attempt to explain the ozone error using the error of selected variables through linear regression analysis achieves very low R2. I would suggest to remove any inference built from this linear model (as it fails to explain the variability of ozone error) and remove figures 17 and 18. Alternatively, the authors can augment the statistical model with the errors from more explanatory variables.
  The main idea here is to introduce a diagnostic technique that combines the usual regression with confidence estimates and to determine the relative importance of the regressors. We believe that not only the dependence between the ozone error and that of the regressors is informative, but also the lack

of it can be instructive in some instances, especially when the variable of interest (the error of ozone) depends (and/or is strictly related) on a number of other fields which, in turn, depend on the location and time/season. For example, this analysis has allowed us to frame the error of ozone due the error of the precursors (NO, NO_2), and thus to reinforce our point about the daily dynamics being one of the major source of error. The suggestion of adding new regressors is indeed very valid and will be exploited in future, but we have already used all the fields for which observational data were made available in AQMEII. R^2 is an estimate of the explained variability (ratio of the variance explained by regressors and total variance); its square root, R, gives a sense of the overall associativity (linear dependence). Therefore, R^2 = 0.3-0.45 (as in most of the cases of Figures 17 and 18) corresponds to a correlation coefficient of approximately 0.5-0.7, which might not be considered as 'small' in an absolute sense.
The revised figures 17 and 18 now present the continent-wide network average (to account for co-located stations, as requested by reviewer #3), thus only one panel per continent.

Technical Comments:

- [L175] Replace 'Where' with 'where' Done
- [L370-372] Please expand or rephrase Done
- Figures: please increase the font size in the majority of figures. The majority of the figures have been revised to improve readability