

## ***Interactive comment on “An unusual way to validate regional chemistry-transport models” by Lauren Menut et al.***

### **Anonymous Referee #2**

Received and published: 19 October 2016

This work addresses the important issue of the validation of chemistry transport models. The authors present a new methodology in which the traditional approach consisting of comparing measurements with model results for a given time period is extended to comparisons of the same model results with measurements from other years. The authors develop then a specific indicator on this basis that allows discriminating results that are good for the good reason from those that are good only because of highly persistent pattern present in the observations from year to year. While the proposed methodology is original and has a potential to complement the traditional approach, the authors remain unfortunately superficial and qualitative in their way of presenting and applying this methodology. As a consequence, the proposed examples are qualitative as well and are not helpful. Finally, the document is poorly written: (1) English would need revisions throughout the whole document and (2) many sections would need to

C1

be re-written (some suggestions are proposed below).

Major points:

1) The authors mention Solazzo and Galmarini (2016) for their decomposition of the error but they finally focus on the correlation only. As noted by these two Authors but also by many others (the referencing to other works relating to model evaluation should be improved), it is important to look at all three possible source of errors because focusing on the only correlation may lead to the wrong conclusions (see comments below). I'm wondering why the Authors make this choice as the proposed methodology could easily be developed for other indicators that are more representative of the overall model performance (e.g. MSE).

2) The approach proposed by the Authors remains qualitative and the interpretation depend on the setting of an arbitrary threshold (e.g. MYV=0.3 in Figure 3). Throughout the text, the Authors make qualitative judgements (0.6 is good, 0.5 is poor... ). This limits the usefulness of the proposed methodology as we never know what a good value of the indicator is. I do not understand this limitation as it would seem relatively straightforward to calculate a value of the MYV indicator in a similar way but on the only basis of measurements. This observation-based MYV value could then serve as the threshold beyond which model results would be considered good enough.

3) The document is poorly written. Many sections are unclear and lack sufficient details to be understood. Some suggestions are provided below but the whole document should be thoroughly revised.

Minor points:

- 1) P1, l1: The title is not very representative of the work
- 2) P1, l3: “and by natural” → “and natural”
- 3) P1, l19: the transport

C2

- 4) P1, l20: or from the QAERONET
- 5) P2, l1: can be
- 6) P2, l2-3: sentence to be revised
- 7) P2, l4: "spatial representativeness" → "spatial representativeness of the monitoring stations". In addition, this concept is mentioned for the first time and should be defined. Finally, I do not get the added value of mentioning this here.
- 8) P2, l5: "to isolate problems intrinsic to the models,". This is unclear and should be re-phrased
- 9) P2, l6: "relevant": which ones?
- 10) P2, l7: "but often with huge" → "but often require important"
- 11) P2, l8: references should be within brackets
- 12) P2, l15-17 & l18-20: if the authors cite these works, they should explain in a little bit more detail their main aspects and why these are important in the context of their work. All these references are introduced independently from the scope of the work. For example on l18, what is the decomposition about? L17, what did Rea et al. find that is relevant for this work. . .
- 13) P2, l18: scores is often misused in the text. Sometimes as real score, some times meant as correlation. I guess the authors here refer to indicators.
- 14) P2, l23: "we apply these scores to a model simulation" is unclear. I do not understand how to apply a score to a model simulation. Please check all occurrences of "scores" and check relevance.
- 15) P2, l27: provide
- 16) P2, l29: spatial representativeness is not yet defined. Is special representativeness really assessed by this method? I do not believe so (see following comments)

### C3

- 17) P2, l33: Score meant as indicator?
- 18) P3, figure 1: I do not believe this figure helps understanding. The proposed methodology is quite universal and does not require to enter these details
- 19) P3, l7: forcings
- 20) P3, l9-23: these lines are not necessary to the methodology and application
- 21) P4, l4: unclear
- 22) P4, l9: for → in
- 23) P4, l12: variable (Table 1)
- 24) P4, l16: and during → for
- 25) P4, l21: take the same day for another → to re-phrase
- 26) P5, l4: why is correlation the more appropriate metric. Why couldn't we say the same for the bias, for example?
- 27) P5, l5: What is a usual correlation score? A correlation is a correlation and a score a score!
- 28) P5, l11-12: I disagree with the authors. A good correlation score does not indicate that the resolution is adequate, transport is adequate . . . Correlation could be 1 while keeping a huge bias due to a too coarse resolution.
- 29) P5, l16: "particularly": why?
- 30) P5, l20: which differences? Between what?
- 31) P6, l5: why should it be larger than unity?
- 32) P6, l5-6: These lines are totally unclear and should be re-phrased
- 33) P6, l7: have → has

### C4

- 34) P6, l7: why do we want that a good score...": although it may appear straightforward, please give a few words of explanation.
- 35) P6, l9: What is an academic value of the score, what is the score meaning here?
- 36) P6, l10: absolute score but also variable: unclear
- 37) P6, l9-15: this all paragraph is unclear and should be rewritten
- 38) P6, l18-19: 5 times scores in these sentences!
- 39) Figure 3 and Figure 6 seems to be inconsistent in terms of X axis labeling.
- 40) P7, l1: from Figure 3
- 41) P7, l1: we can consider that
- 42) P7, l1-2: This means that all conclusions will remain subjective because of this arbitrarily fixed delta parameter. I believe that a measurement based threshold value for delta can be fixed, withdrawing this arbitrary aspect (see major comment above).
- 43) P7, l6: done → calculated
- 44) P7, l6: MYV scores
- 45) P7, l12: vary a lot → vary significantly
- 46) P7, l13: is challenging because
- 47) P7, l13: again spatial representativeness needs to be defined
- 48) P7, l17: "The spatial correlation is good for all years". I do not understand which arguments the Authors use to state that the score is good. If the spatial pattern is easy to reproduce, it could well be that a correlation of 0.7 should be considered as bad. This seems to be confirmed by the next sentence: "the model reproduces fairly well a spatial pattern observed every year". One way forward is to calculate the correlations on the only basis of measurements to get some indicative threshold of what is good or

C5

not.

- 49) P8, l2: Are we sure this is for the good reasons?
- 50) P8, l6: "This species is secondary" seems to contradict p7, l12.
- 51) P8, l6,7: I do not agree that a good score for correlation is indicating a good transport, photochemistry... Correlation is indeed only one of the indicators to assess model performances and it only provides a partial vision of model performances. Correlation could be perfect even with a very large bias.
- 52) P8, l8: low → coarse
- 53) P8, l8: less good → worse
- 54) "Its spatial extent of its representativeness": totally unclear, this should be re-phrased
- 55) P8, l18: "The scores": The correlations are calculated, not the scores which are the correlation values
- 56) P8, l20: "each score type". I do not understand what the Authors mean.
- 57) P8, l20: "Results are presented in Table 3. These results..." → Results (Table 3) are discussed...
- 58) P8, l24: why only?
- 59) P8, l24: Which arguments are used to state that the spatial correlation is not correct?
- 60) P8, l24: for one year → from one year
- 61) P8, l26, 27 and 28: "very good spatial", "less good", "well retrieved". The Authors should explain how they come to these statements.
- 62) P8, l31: A few words to explain what the AOD and ANG are would be helpful

C6

- 63) Figure 4 caption: Should include explanations of the two curves represented
- 64) P10, l9,10,11: Again I do not agree with these conclusions which cannot be drawn from the only correlation values.
- 65) P11, l19-20: this sentence is unclear
- 66) P12, l29: dued → due

---

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016.