

1 **Author reply to the reviewers of “Air Control Toolbox (ACT\_v1.0): a machine**  
2 **learning flexible surrogate model to explore mitigation scenarios in air quality**  
3 **forecasts.” Submitted to Geos. Model Dev. By Augustin Colette, Laurence**  
4 **Rouil, Frédérik Meleux, Vincent Lemaire, Blandine Raux**

5  
6 We are deeply grateful to the three reviewers who provided constructive critics and excellent  
7 reflexions to improve our manuscript, yet supporting the approach undertaken in this development.  
8 We tried our best to address all concerns, with detailed answer in this document, and in track change  
9 in the original manuscript.

10 In this response, the reviewer comments are highlighted in bold, author responses in plain text, and  
11 modifications to the manuscript indented in track changes with page and line numbering referring to  
12 the track change version of the revised manuscript.

13 **Community Comment #1 (<https://doi.org/10.5194/gmd-2020-433-CC1>)**

14 **ABTRACT. In the abstract, the authors should better explain the limitations of their approach.**  
15 **Mainly, the fact that, in my understanding, scenarios are related to EU emission reductions, applying**  
16 **constant sectoral reductions everywhere over the domain. These are quite important assumptions,**  
17 **that limit the applicability of the tool to specific type of studies. Authors should better stress this**  
18 **points.**

19 This comment is totally relevant. And the reviewer is right that is was not so clear in the abstract. We  
20 have made it clearer with the following changes.

21 In the abstract: P1L26 of the revised manuscript

22 The selected approach makes ACT the first air quality surrogate model capable to capture non-linearities  
23 in atmospheric chemistry response. Existing air quality surrogate models generally rely on a linearity  
24 assumption over a given range of emission reductions, which often limits their applicability to annual  
25 indicators. Such a structure makes ACT especially relevant to understand the main drivers of Various  
26 applications of the toolbox are presented for air pollution quality-episode analysis. This feature is a strong  
27 asset of this innovative tool which makes it also relevant for, source apportionment, and chemical regime  
28 analysis. This breakthrough was only possible by assuming uniform and constant emission reductions for  
29 the four targeted activity sectors. This version of the tool is therefore not suited to investigate short term  
30 mitigation measures or spatially varying emission reductions.

31 **INTRODUCTION. The sentence “The two most widespread applications of atmospheric chemistry**  
32 **modelling are (i) short term air quality forecasting, and (ii) long term analysis of mitigation strategies.**  
33 **We introduce here the first toolbox able to address both issues at once,” is for me misleading. As**  
34 **long-term mitigation strategies is more for structural measures for which a forecasting model is not**  
35 **useful. Or better, you can use a forecasting model, but it is more useful to check the change of**  
36 **concentrations in the long-term, more than the day to day change. Furthermore, when designing**  
37 **mitigation scenarios, is quite uncommon to have constant reductions over the whole domain. Please**  
38 **better specify these limitations.**

39 A clarification of the fact that only long-term mitigation strategies are targeted has been added in the  
40 first sentence of the introduction, P2L12 of the revised manuscript

1 We introduce here the first toolbox able to address both issues at once, so that the user can explore the  
2 benefit of a long term any-emission reduction control strategy for the current-day air quality forecast.

3 The assumption that emission reductions are homogeneous and constant were already clearly  
4 highlighted later in the introduction (P3L19-20 of the submitted manuscript). But we further expanded  
5 that paragraph to highlight that we do not intend to “check the change of concentrations in the long-  
6 term” as rightfully pointed out by the reviewer. It is indeed the day-to-day change which is our focus.  
7 When it comes to decide on which of them is “more useful”, we would leave it up to potential users.  
8 The position of the authors is that it remains interesting to know the drivers of the daily variability in  
9 air pollution. We can make an analogy in the field of climate science where some studies intend to  
10 quantify the role of climate change in current weather extremes. This is indeed different than assessing  
11 the benefit of long-term strategies to mitigate climate change. Understanding the difference is  
12 important but deciding on which of them is more useful is a matter of personal opinion. Nevertheless,  
13 the reviewer is right that such assumption must be well explained to the reader, and we hope that the  
14 revision proposed below addresses this point.

15 In the introduction, P4L16 of the revised manuscript:

16 The only two simplifications limiting the range of application of ACT are that emission reductions are  
17 assumed to apply (i) over the long term (~~so that it is not possible to investigate an emergency mitigation~~  
18 ~~measures~~) and (ii) over the whole modelling domain uniformly. ACT is therefore not suited to assess the  
19 benefit of emergency measures to mitigate air pollution episodes. It also not designed to assess the impact  
20 on long term exposure to air pollution. The purpose of the tool is rather to assess the main activity sectors  
21 driving the day to day air pollution variability. Several diagnostics are proposed to achieve this by  
22 providing the source allocation and chemical regimes at various receptor locations, as well as mapping  
23 the benefit of reducing emissions over the whole modelling domain. If we take the analogy with climate  
24 change, the scope of ACT is analogous to attribution studies where one intends to quantify the role of  
25 climate change in the current meteorological situation. Which is different than assessing the benefit of  
26 greenhouse gas strategies to mitigate long term climate change.

27 A similar note was also added in the conclusion, P24L19 of the revised manuscript:

28 To our knowledge, this model is the first surrogate, or emulator, able to cover short time scale for air  
29 pollution studies. This development was only made possible by assuming uniform and constant emission  
30 reductions for the four targeted activity sectors. The fact that emission reductions are assumed to be  
31 applied over the long term makes ACT not suited to assess the benefit of emergency measures to mitigate  
32 air pollution episodes. The purpose of the tool is rather to assess the main activity sectors driving the day  
33 to day air pollution variability.

34 **SECTION 3. The paper in section 3 is much too long and full of details. I would propose to restructure**  
35 **the work, moving to Supplementary Material the technicalities of the work, and keeping only the**  
36 **most relevant parts of the paper in the ‘Main’ part.**

37  
38 Section 3 of the paper constitutes the main body of the design of the ACT tool. In the submitted  
39 manuscript it is 7.5 pages long out of 22 (without references and figures) which can also be considered  
40 an appropriate level of detail. We believe that a model description paper in GMD is well suited for such  
41 a methodological description, unlike other journals where methods are moved to the supplementary  
42 materials. But we can reconsider this position if the editor thinks fit.

2  
3 **Abstract and introduction: Overall, the paper is well written but the authors should introduce more**  
4 **clearly the objective of their study and the general form of the tool they aimed to build. While**  
5 **reading the paper, it took me some time to understand precisely which type of tool they were**  
6 **developing. One reason may be that in the abstract and introduction, the authors are often talking**  
7 **about a “surrogate model” to tackle the complexity of geophysical chemistry-transport models**  
8 **(CTM). I was a bit confused at the beginning since this tends to suggest that they want to build a**  
9 **CTM surrogate model (able to predict concentrations based on emissions) while actually they want**  
10 **to develop an emission scenario surrogate model (able to predict the concentration changes in**  
11 **response to emission changes). Although some sentences indeed refer to the later, some others refer**  
12 **to the former, which maintains the confusion (e.g., P2L6 : “an emulator, or surrogate model, of a**  
13 **comprehensive air quality model”). While reading the paper, we progressively understand what the**  
14 **authors were referring to before, but the first part of the paper could be clarified so that things are**  
15 **clear earlier in the text. I would personally understand better and easier the goal of their study by**  
16 **introducing it in such a way :**

17 **The overall objective of the authors is to build a tool for conducting flexible emission scenario**  
18 **analysis on an every-day basis, for exploring the impact of different mitigation strategies on the**  
19 **current forecast. More specifically, they want the tool to estimate the impact of a given set of**  
20 **emission reductions (in this case, looking at the options available for the user on the ACT web**  
21 **interface, 21 emission scenarios of emission reduction bins of 5%, namely 0%, -5%, -10%..., -95%, -**  
22 **100%) applied to a given set of emission sectors (in this case, 4 emission sectors). Simulating**  
23 **explicitly all possible combinations would require a total of  $21^4=194,481$  emission scenario**  
24 **simulations (or  $101^4=104,060,401$  simulations with emission reduction bins of 1%), which is**  
25 **obviously not feasible for computational reasons. This motivates the need for developing, in each**  
26 **cell of the domain and for each day of forecast, (1) a regression model able to estimate the pollutant**  
27 **concentration responses to different emission reductions based on (2) a reasonable number of**  
28 **emission scenario simulations that could be run in parallel to the reference CTM simulation on an**  
29 **every-day basis.**

30 **Then, this naturally justifies the type of regression model chosen. Indeed, although more complex**  
31 **regression models can eventually achieve better accuracy, they require larger training datasets (in**  
32 **this case, more emission scenario simulations). As a compromise between these two aspects, this**  
33 **motivates the choice of multivariate second-order polynomial models with interaction terms able to**  
34 **capture some of the non-linearities existing between concentration changes and emission change**  
35 **(at least in some emission sectors) while remaining simple enough, in the sense that only a few**  
36 **coefficients need to be fitted, which allows relying on a limited number of emission scenario**  
37 **simulations.**

38  
39 We are sorry if the reviewer found it difficult to understand the scope of the work in the abstract and  
40 introduction. We very much appreciate the suggestions for improvement, which helped in improving  
41 the abstract and introduction.

42  
43 In the abstract: P1L16 of the revised manuscript

44  
45 As such we take the best of the physical and chemical complexity of CTMs, operated on high performance  
46 computers for the everyday forecast, but we approximate a simplified response function that can be  
47 operated through a website to emulate the ~~main~~-sensitivity~~ies~~ of the atmospheric system to anthropogenic  
48 emission changes for a given day and location.

49  
50 In the introduction: P3L22 of the revised manuscript

1 The overall objective is to offer the users a high degree of flexibility for any  
2 user to explore any their own mitigation scenario through a web interface. We are targeting four activity  
3 sectors and for each of them the available emission reduction should cover the whole 0 to 100% range,  
4 for instance by 5% increments. Using an explicit approach, this would imply  $21^4$  or 194,481 chemistry  
5 transport simulations which is obviously not feasible for computational reasons. The whole point of the  
6 present paper is therefore to design a methodology to reduce this number down to a reasonable number  
7 of simulations which can be performed on a daily basis and used as a training set to calibrate a simple  
8 response model.

9 The selected architecture of the response model is a polynomial function whose coefficients can be  
10 explicitly estimated from a limited number of simulations. Fitting more complex regression models could  
11 achieve better accuracy, but only at the cost of a larger training set. We will see that a quadrivariate (four  
12 activity sector) second order polynomial can be well constrained with only a dozen of chemistry transport  
13 simulations, which is quite acceptable with regards to the operational constraints of every-day production  
14 in an air quality forecasting system.

15 With this approach, the model is simplified down to the level of a polynomial function. Nevertheless, all  
16 the complexity of the atmospheric chemistry response to incremental emission changes is embedded in  
17 this a polynomial surrogate which is derived fitted every day from based on a subset (typically 10 to 15)  
18 of chemistry-transport simulations.

19 The use of a complete chemistry-transport model to build the training simulations allows, thereby relying  
20 on automated machine learning approach. As a result, the surrogate polynomial model accountings for all  
21 the important processes bearing upon the forecasted air quality, including long-range transport and  
22 chemistry. And the surrogate model is able capture some of the non-linearities existing between  
23 concentration and emission changes while remaining simple enough, in the sense that only a few  
24 coefficients need to be resolved.

25  
26 **Section3 : I agree with the first reviewer that section 3 could be shorter. Also, it is not very clear to**  
27 **which extent the step-by-step approach - (1) univariate first/second/third order model, (2) bivariate**  
28 **second order model with interaction terms, (3) quadrivariate first/second order model with**  
29 **interaction terms - is followed only for pedagogical purposes or if (at least part of) the methodology**  
30 **is really sequential (it seems so). Please clarify.**

31  
32 As replied to CC1, we regret this impression of a too long section 3 (which is only 7.5p out of 22). We  
33 leave it up for the editor guidance to move part of it to the supplementary material, but we also  
34 understood that it is the very scope of GMD to publish such details, while other journals require to  
35 keep the methodological details in the annex.

36  
37 This sequential approach has indeed some pedagogical advantages. But the reviewer is right in pointing  
38 out that the design of the method indeed follows this series of steps. That is precisely why we kept this  
39 structure in the paper as we believe that the same sequence should be followed when implementing  
40 the method in a different context. The ACT tool presented here covers Europe at 0.25 degree  
41 resolution. But one may want to replicate the approach in other areas, or at finer spatial resolution.  
42 Nothing guarantees that the same model structure would hold in such cases. And we advise that the  
43 same series of steps are followed.

44  
45 Section 3.3 P15L4 of the revised manuscript

46  
47 The methodology followed in Sections **Erreur ! Source du renvoi introuvable.** and **Erreur ! Source**  
48 **du renvoi introuvable.** consists in selecting first the optimal structure for univariate models, before  
49 investigating bivariate models including interactions terms. Such a step-by-step approach allows a clear  
50 introduction of the methodology. This presentation has a clear pedagogical advantage. It is also relevant  
51 for a potential development of a similar approach in a different context. One could consider developing  
52 an ACT tool over a different region, or at higher spatial resolution. But, in that case, nothing guarantees

1 that the same model structure would be selected, in particular if non-linearities affect different activity  
2 sectors.

3 However, such a sequential approach carries a risk of not selecting the optimal structure, as pointed out  
4 for stepwise regression approaches. Indeed, with a sequential method~~by doing so~~, we assume that the  
5 optimal structure and training scenario remains valid when including interactions whereas there is a  
6 possibility that the addition of an interaction term could change the selection of univariate terms.

7  
8 **Section 3: If I understand correctly, pairs of emission sectors with most important interactions are**  
9 **chosen in section 3.2 based on different two-variable polynomial regression models, but how this**  
10 **ensures that it remains then the most appropriate choice in the four-variable polynomial models**  
11 **developed in section 3.3? The authors acknowledge that “by doing [such a step-by-step approach],**  
12 **[they] assume that the optimal structure and training scenario remains valid when including**  
13 **interactions whereas there is a possibility that the addition of an interaction term could change the**  
14 **selection of univariate terms.” However, in the type of model proposed in section 3.3, the authors**  
15 **keep only the interaction terms selected before. Also, to test this final model, they retain only the**  
16 **subset of emission scenario selected before. These different methodological aspects should be**  
17 **clarified.**

18  
19 Despite interesting from a pedagogical point of view, the chosen sequential approach carries a risk of  
20 deviation in the optimum also identified for stepwise regression. That is why, after having identified  
21 an optimum with a sequential method, we explore again all possible combinations in section 3. With  
22 this systematic combination of all models, we show that the model selected with the sequential  
23 approach turns out to be the optimum for ozone, but not for PM<sub>10</sub> (by only a small margin). All the  
24 relevant pairs of interaction terms are kept. Only those involving the RH sector are excluding because  
25 it has been demonstrated to be well approximated with a linear fit.

26  
27 Section 3.3 P15L23 of the revised manuscript

28 Such a model requires two training scenarios for AGR, IND, TRA, one for RH and one for each of the  
29 three interaction terms. With this approach all possible 2-term interactions are indeed taken into account  
30 as only those involving RH are excluded because we have demonstrated earlier that this factor could be  
31 well approximated with a linear relationship (and therefore irrelevant for second order interactions).

32 **I wonder to which extent not considering any emission scenario with reductions applied to 3 or 4**  
33 **sectors simultaneously (besides the scenario with all 4 sectors reduced by 100%), for the training or**  
34 **at least the testing is a strong limitation in this study. This should be discussed (and ideally tested).**

35  
36 This comment is very relevant, and one could indeed consider further increasing the refinement in the  
37 surrogate model. The only reason not to engage in that direction is a trade-off between computational  
38 cost of the training scenario and performances of the surrogate.

39  
40 A discussion has been added Section 3.4 P17L7 of the revised manuscript:

41 Further increasing the degree of the polynomial would certainly improve the quality of the surrogate  
42 model. The only two reasons not to engage in that direction are: (i) avoid increasing the computation  
43 burden with more training scenario and (ii) considering that the performance achieved at order 2 are  
44 already very satisfactory. The only higher order interaction scenario removing emissions from all 4 sectors  
45 is designed as a closure to avoid any potential negative concentrations.

46 **Additional comments :**

47 **P1L13 : “flexibility” : please be more specific on which type of flexibility you are talking about (here,**  
48 **flexibility on the emission forcing)**

49 Abstract P1L13 of the revised manuscript:

1 However, their complexity prohibits offering a high level of flexibility in the tested emission reductions.

2 **P4L11 : “Note however that those periods do not constitute a specific training period for the**  
3 **surrogate model which is intended to be re-fitted to new CTM simulations every forecasted day in**  
4 **an automated machine learning approach.” Do the authors mean that all regression models in this**  
5 **study are trained only based on the current forecast day? Please clarify. In total for this study, the**  
6 **authors have performed 46 CHIMERE simulations for each of the 4 months mentioned in Section 2.1,**  
7 **is it correct?**

8 A clarification has been added in Section 2.1, P5L17 of the revised manuscript:

9 Note however that those periods do not constitute a specific training period for the surrogate model which  
10 is intended to be re-fitted to new CTM simulations every forecasted day in an automated machine learning  
11 approach. Those episodes are only selected to identify the best design for the surrogate and assess its  
12 performances in reproducing the full CTM. We will see in Section Erreur ! Source du renvoi  
13 introuvable. that in total 46 simulations were required to identify the optimum surrogate model structure  
14 and demonstrate its performances. All these 46 simulations covered the 4 months selected in the years  
15 2015 to 2017 to capture a variety of air pollution episodes. But once the structure of the surrogate model  
16 is identified, the model itself is intended to be calibrated automatically on the basis of the day to day  
17 forecast.

18 **P4L25 : “operational analyses of the IFS” : not IFS forecasts?**

19 Indeed for this development phase, IFS analyses were used, but in operational setup, ACT is fitted on  
20 CHIMERE simulations driven by IFS forecasts

21 **P5L20 : “we conclude that the surrogate model will be at most a third order polynomial, less if**  
22 **interaction terms are accounted for.” Please clarify that this is motivated by the number of**  
23 **coefficients to be fitted.**

24 Section 2.3, P7L1 of the revised manuscript:

25 Considering the goal to cover four activity sectors, and the operational constrain to compute only with  
26 about 10 to 15 training scenarios every day, we can derive a limit in the number of coefficients which  
27 we can estimate. We conclude that the surrogate model can will be at most a third order polynomial, or  
28 even less if interaction terms are accounted for.

29 **P6L4 : “emissions are applied uniformly over Europe” should be “emission reductions...”**

30 Section 2.4, P7L14 of the revised manuscript:

31 In all cases, emissions reductions are applied uniformly over Europe and for all chemical species.

32 **P7L16 : “both dates” : which dates are you referring to?**

33 Section 3.1.1, P9L2 of the revised manuscript:

34 The residential heating (RH) sector contributes mainly with primary PM<sub>10</sub> emissions or organic aerosol  
35 precursors, and its response is the closest to linearity for 20<sup>th</sup> March 2015 and 1<sup>st</sup> December 2016 both  
36 dates and the three selected cities.

37 **P9L1 : “testing and validation” should be “training and testing”**

38 Section 3.1.1, P11L1 of the revised manuscript:

39 The corresponding list of scenarios available for training and testing and validation are summarized in  
40 **Erreur ! Source du renvoi introuvable.**, there are 5, 10, and 10 combinations for the linear, quadratic  
41 and cubic forms, respectively.

1 **FIG3 : Please describe in the legend the meaning of the numbers into brackets (I guess they**  
2 **correspond to the minimum and maximum error over the domain?)**

3 The legend of Figure 3 has been changed following a comment from Anonymous Reviewer #2.

4 **P10L1 : “perform” should be “performs”; “largest emission reduction” should be “largest emission**  
5 **reduction”**

6 The first change was taken into account. I could not find the issue with the second point

7 Section 3.1.2, P11L24 of the revised manuscript:

8 The third order polynomial performs best, except the last model which uses only the largest emission  
9 reduction and is therefore too weakly constrained for the lower range of reductions

10 **P11L6 : “yield” should be “yields”**

11 Section 3.1.2, P13L4 of the revised manuscript:

12 On the contrary, for Industry and Traffic, we opt for a quadratic model (using 60% and 100% reductions,  
13 index 14 in the x-axis of **Erreur ! Source du renvoi introuvable.**) which yields median errors below  
14 0.1% (0.099 and 0.028%, for IND and TRA, respectively) and the gain in term of median error is a factor  
15 3-4 compared to the linear forms.

16 **Sect. 3.1 : Results are discussed for PM10 and briefly mentioned for O3max but not for the other**  
17 **pollutants**

18 This is because the same reasoning applies for other pollutants and metrics, for which only overall  
19 performances are presented in Section 3.4

20 Section 3.4, P16L27 of the revised manuscript:

21 More specifically, the single highest error over the 864 248 grid points and days considered is 7.5%, 7.9%,  
22 4.8%, 2.8% and 2.9% for PM10, PM2.5, NO2, O3max, O3avg

23 **Table 2 : a comma is missing between IND30 and IND60**

24 Added in Table 2

25 **P12L10 : It seems to me the authors are actually talking about lines 3 and 4 where no interactions**  
26 **are included in the training set, please correct if needed.**

27 Section 3.2, P14L28 of the revised manuscript:

28 First a bivariate quadratic model without interactions is trained with the 30% and 60% reduction levels  
29 and tested against corresponding interaction scenarios. Taking the example of agriculture & industry, we  
30 would have the two training and testing configurations in lines 31 and 42 of **Erreur ! Source du renvoi**  
31 **introuvable..**

32 **All formulas : In order to make formulas easier to read (especially the longest ones), I suggest to**  
33 **introduce another variable such as DELTA^sector=EPSILON^sector-EPSILON^ref.**

34 This notation was introduced in all relevant equations in Section 3.1.2, 3.2, and 3.3 (P10L10, P14L1,  
35 P15L15 of the revised manuscript)

36 
$$C_{i,j} - C_{i,j}^{ref} = \alpha_{i,j} \cdot \left( \delta_{i,j}^{sec} \cdot \epsilon_{i,j} - \epsilon_{i,j}^{ref} \right) + \beta_{i,j} \cdot \left( \delta_{i,j}^{sec} \cdot \epsilon_{i,j} - \epsilon_{i,j}^{ref} \right)^2 + \gamma_{i,j} \cdot \left( \delta_{i,j}^{sec} \cdot \epsilon_{i,j} - \epsilon_{i,j}^{ref} \right)^3$$

37 
$$C^{agr,ind} - C^{ref} = \alpha^{agr} \cdot \delta^{agr} + \beta^{agr} \cdot (\delta^{agr})^2 + \alpha^{ind} \cdot (\delta^{ind}) + \beta^{ind} \cdot (\delta^{ind})^2 + \gamma \cdot (\delta^{agr}) \cdot (\delta^{ind})$$

$$\begin{aligned}
& C^{agr,ind,rh,tra} - C^{ref} \\
& = \alpha^{agr} \cdot (\delta^{agr}) + \beta^{agr} \cdot (\delta^{agr})^2 + \alpha^{ind} \cdot (\delta^{ind}) + \beta^{ind} \cdot (\delta^{ind})^2 + \alpha^{rh} \cdot (\delta^{rh}) + \alpha^{tra} \\
& \cdot (\delta^{tra}) + \beta^{tra} \cdot (\delta^{tra})^2 + \gamma^{agr,ind} \cdot (\delta^{agr}) \cdot (\delta^{ind}) + \gamma^{tra,agr} \cdot (\delta^{tra}) \cdot (\delta^{agr}) \\
& + \gamma^{tra,ind} \cdot (\delta^{tra}) \cdot (\delta^{ind})
\end{aligned}$$

5 **P14L13 : “This final model is tested the” should be “This final model is tested against the”**

6 Section 3.4, P16L23 of the revised manuscript:

7 This final model is tested against the 34 available scenarios not used in the training.

8 **P17L6 : “ammonia emissions December is” should be “ammonia emissions in December are”**

9 Section 4.2.2, P19L28 of the revised manuscript:

10 Given the strong seasonality of fertilizer spreading, ammonia emissions in December are ~~is~~ likely due  
11 mainly to livestock emissions.

12 **P18L4 : “the allocation we display here” : where?**

13 Section 5, P20L24 of the revised manuscript:

14 Unlike source apportionment, the allocation we introduce ~~display~~ here indeed shows the reduction of  
15 concentration that can be achieved by removing totally the emissions of an activity sectors.

16 **P18L14 : “assess” should be “assessed”**

17 Section 5, P21L7 of the revised manuscript:

18 Here, the contribution of each activity sector is assessed by setting its emissions to zero (referred to as  
19 top-down brute force method in (Clappier et al., 2017a)).

20 **ACT web interface : Why only O3 is proposed and not O3max or O3avg? Also, the source allocation  
21 model is useful for analysing pollution episodes and I would suggest to include it on the ACT web  
22 interface in the future.**

23 Adding O3avg as well as source allocation and chemical regimes are currently being considered for  
24 inclusion in the web tool. Given that they are also sometimes more difficult to understand for a non-  
25 specialised audience, the publication of the present paper was an important prerequisite prior to such  
26 a public dissemination.

27



2  
3  
4 This paper addresses an important topic, is interesting and well written. The authors develop a  
5 surrogate model for a full CTM to investigate ambient pollution changes in response to sectoral  
6 emission reductions. Currently the application is limited to episodes and uniform emission changes  
7 over Europe which, as also remarked by the other anonymous referee, limits the practical applicability  
8 for policy analysis, but the authors also emphasize that they are planning to extend the tool in this  
9 direction. It will be interesting to follow this development.

10 I have a few comments which should be addressed, but these are mostly clarifications and should not  
11 require major changes before publication.

12  
13 General comments

14  
15 **1. I confess I am not an expert for machine learning but I was wondering whether the title is really**  
16 **appropriate. The toolbox as described in this paper uses a parametric regression (first and second**  
17 **order with interaction terms) to approximate the behaviour of a full CTM, and while I think the**  
18 **methodology is sound, I'm not sure if this can be really termed machine learning**

19  
20 The reviewer is right that we used a probably too broad definition of machine learning. The toolbox is  
21 designed to be calibrated automatically, but the actual structure of the surrogate is designed manually  
22 and indeed relies on a parametric model. In using that term in the paper, we followed a general  
23 tendency to include any statistical methods under the terminology of machine learning, which is  
24 perhaps not justified, so that we removed these instances in the title (P1L1), abstract (P1L14), and  
25 section 2.1 (P5L15).

26  
27 **2. I struggled a bit to understand how the regression is set up. When looking at Eq 1, one could get**  
28 **the impression that everything is known, and the equation can simply be solved with a unique**  
29 **solution in each grid cell for the parameters alpha, beta, gamma (if  $C_{i,j}$  and  $\epsilon$  are scalar values**  
30 **averaged for the time period of interest) and no regression is needed if the right number of training**  
31 **scenarios is used (2 for a linear fit, 3 for quadratic etc). Is this what is done, or is there a regression**  
32 **involved in the sense of solving an overdetermined equation system and estimating optimal alpha,**  
33 **beta, gammas? Also, should Eq 1 not have a sectoral dimension index in alpha, beta, gamma and**  
34 **epsilons?**

35  
36 We apologize that the term “regression” was misleading. The system is not overdetermined: because  
37 of computational constrains we don't have the luxury to run unnecessary simulations with the full  
38 chemistry transport model.

39  
40 This has been made clearer by avoiding referring to “fit” or “regression” in more than 20 instances  
41 throughout the paper.

42  
43 The sectoral dimension was lacking in Eq 1, it has now been added, P10L10 of the revised manuscript:

44  
45 
$$C_{i,j} - C_{i,j}^{ref} = \alpha_{i,j} \cdot (\delta_{i,j}^{sec} \epsilon_{i,j} - \epsilon_{i,j}^{ref}) + \beta_{i,j} \cdot (\delta_{i,j}^{sec} \epsilon_{i,j} - \epsilon_{i,j}^{ref})^2 + \gamma_{i,j} \cdot (\delta_{i,j}^{sec} \epsilon_{i,j} - \epsilon_{i,j}^{ref})^3$$

46 **3. Besides the limitation that emission reductions which can be explored with this tool are uniform**  
47 **across Europe, also they are specified as reduction of emissions by sector since the coefficients are**  
48 **expressed by sector and not by pollutant, meaning that all pollutants emitted from this sector are**  
49 **reduced proportionally. In practice, emission control policies may reduce different pollutants from**

1 the same SNAP sector to a different extent, which cannot be simulated with the current setup. It  
2 would be good to mention this somewhere.

3  
4 We fully agree with this comment, but would like to emphasize that this is complementary with other  
5 approaches, but not a limitation per se. Otherwise one should consider that surrogate models targeting  
6 individual species are also limited in the fact that they do not provide the flexibility to assess activity  
7 sectors. The following sentences were added P7L14 of the revised manuscript:

8  
9 In all cases, emissions reductions are applied uniformly over Europe. It is also important to stress that  
10 and for all chemical species for a given activity sector are reduced by the same amount. This is different  
11 and complementary with the approach chosen for instance in the EMEP source receptor matrices (Amann  
12 et al., 2008) or the SHERPA tool (Pisoni et al., 2017).

13  
14 **P 5, L 6: The number of training simulations should be constrained by more fundamental**  
15 **principles than the availability of computing time?**

16  
17 The tool is specifically designed to be implemented in an operational daily air quality forecasting  
18 system. Therefore computing time constrains are indeed a critical limitation to be taken into account  
19 in its design.

20  
21 **P 7 – I would emphasise a bit more the complete reverse response for TRA reductions wrt O3 avg vs**  
22 **peaks. In this context it would be good to mention what is assumed regarding NO/NOx ratio in**  
23 **industrial vs traffic emissions, this may play a role here.**

24  
25 This paragraph has been expanded to emphasise this difference and include a mention to the role of  
26 NO/NOx ratios, cf P9L17 of the revised manuscript:

27 This very distinct sensitivity of O<sub>3</sub> daily maximum and daily averages is due to the titration process, which  
28 affects more strongly night-time, low, ozone concentrations and is therefore less visible on the daily  
29 maximum. In Figure 2, it is a specific feature of the traffic sector, which is not seen for instance for  
30 industrial emissions. It should be noted that the NO<sub>2</sub>/NOx ratio is very different for both sectors. In the  
31 CHIMERE modelling setup used here, 20% of NOx emissions are allocated to NO<sub>2</sub> for the traffic sector,  
32 while for other sources including industry it is only 4.5% of NOx emissions that are constituted of NO<sub>2</sub>.  
33 But this larger share of NOx emissions as NO in the industry sector would rather imply a lower titration  
34 of O<sub>3</sub> by NO in traffic emissions compared to industrial sources. And in summer, NO and NO<sub>2</sub> will even  
35 out very fast at the spatial scales of a regional chemistry transport model. The stronger impact of titration  
36 when reducing traffic emissions in Figure 2 is therefore rather a consequence of the overall larger share  
37 of total NOx emissions compared to other activity sectors.

38  
39 **P10 L 22: I am confused what is actually calculated and shown here. Maybe I'm too long after my**  
40 **last statistics class but I don't exactly understand why a confidence interval is calculated here (for**  
41 **what estimated quantity?) and where the formula comes from. Wouldn't it be more useful to show**  
42 **just the 95th or 97.5th percentile of this error distribution?**

43  
44 We used the default implementation of boxplot displays in R, and we appreciate this comment which  
45 provided us the opportunity to fix an inaccuracy in the paper. After a closer look it appears that these  
46 whiskers are not providing the confidence interval but rather 1.5 the interquartile range. See new  
47 formulation in Section 3.1.2, P12L19 of the revised paper:

48 W<sub>i</sub> and whiskers extend to 1.5 times the interquartile range from the borders of the box, and the extreme  
49 points lying outside of that range are providing the 95% confidence interval (as 1.58 times the  
50 interquartile range divided by the square root of the sample size). The points standing out of the 95%  
51 confidence interval are also provided.

52  
53 **Section 3.4: Is the final model setup that the authors arrive at in Section 3.4 stable or is**  
54 **this re-evaluated for new time periods in the operational tool?**

1 This has been clarified in 3.4, P17L1 of the revised paper:

2 This structure is selected because of its good performances for all relevant ambient air pollutant as  
3 demonstrated here over a range of air pollution episodes. In the operational model, the overall structure  
4 of the model is frozen, while only the coefficients of the surrogate are recomputed every day on the basis  
5 of the current air quality forecast. It should be noted however that any evolution of the model such as  
6 increasing the spatial resolution, implementation over a different geographic area, or including other  
7 activity sector would require revising this structure.

8  
9 **Section 4.2.2: The conclusion that the December episode is mainly driven by agricultural emissions**  
10 **is rather unexpected, I would have expected a higher contribution from residential heating. How**  
11 **well does the CHIMERE model in this configuration perform in reproducing the measured**  
12 **concentrations of these episodes? The authors note that the emission inventory used may**  
13 **underestimate residential sector emissions, which is an important caveat to highlight when drawing**  
14 **conclusions for wintertime episodes.**

15  
16 We agree that it is important to highlight the limitations introduced by the underlying model and data.  
17 This has been further emphasised in Section 5 on Source Apportionment, P21L17 of the revised  
18 manuscript:

19 This underestimation of residential emission has been widely documented as it has a very strong impact  
20 on model performances to capture wintertime peaks (Denier van der Gon et al., 2015). The expected  
21 future improvements in the consistency of reported condensable PM in emission inventories should  
22 improve notably model performances. Until then, this feature provides us an opportunity to highlight that  
23 the diagnostics derived with the surrogate model remains highly sensitive to underlying hypotheses in  
24 emission inventories and chemistry transport modelling.

25  
26 All the following technical corrections were implemented, as noted below:

27  
28 **P 6, L 5. “Emissions”: emission changes / reductions**

29 Done

30  
31 **P9 L21 performance (singular)**

32 Done

33  
34 **P10 L12: in order to reduce the number of performances in this sentence, I suggest to**  
35 **replace the middle one with ‘done’**

36 Done

37  
38 **Fig 3: The legend is messy and the line colors are not distinguishable. I suggest to change the figure**  
39 **layout. For example, unify to one table with scenario numbers like in Fig 4.**

40 Figure 4 has been modified to account for this suggestion. The same modification was applied to  
41 Figures S.3, S.4, and S.5.

42  
43 **Fig 4: color scales are exhausted for some of the figures**

44 This is unfortunately true, but we selected the color scale range specifically to provide a relevant level  
45 of details for the quadratic models which provide the best results

46  
47 **P13 L19 – “We are left already” -> We are still left**

48 Done

49  
50 **Figure 7 caption: raw -> row**

51 Done

52  
53 **P14 L13 – is tested with / against**

- 1 Done
- 2
- 3 **Figure 11 caption: 2015903 ?**
- 4 Done
- 5
- 6 **Figure 12 top right - caption unclear**
- 7 A clarification was added in the main text introducing that figure (Section 4.2.1, P18L14 of the revised
- 8 manuscript)
- 9
- 10 **P16 L 10: NOx from industry and heating**
- 11 Done
- 12
- 13 **P16 L18: in Western Europe**
- 14 Done
- 15