Reply to referee #1:

Issue: The referee suggests that the following statement in line 124 is debatable, since not all reactions have an important dependence upon orientation: "The pre-exponential A-factor represents the rate of molecular collisions with the correct orientation for reaction and the exponential term is the fraction of those collisions with sufficient energy for reaction to occur."

Response: It is our understanding that the experimentally derived A-factor is distinguished from the collision frequency, Z, by the fact that not all collisions can result in a reaction, which is generally considered to be because of steric effects/ orientation. It is also important to point out that for any reaction which has a small or negligible steric effect, then essentially all collisions have the correct orientation, which does not conflict with our current definition.

Action: No proposed changes.

Issue: The referee suggests that there could be a missing sign in the exponential term in the equation in line 142?

Response: We assume that the referee could mean either the *n* parameter or the Ea/R parameter could be negative, the answer to which is no, neither are negative in this case, though they can be negative in other cases.

Action: No proposed changes.

Issue: The meaning of the sentence in line 144 is unclear. We presume this refers to the sentence: "It can be observed that, whether the measurement is absolute or relative, that the vast majority of kinetic studies provide total phenomenological rate coefficients rather than site-specific data"

Action: Sentence changed to read:

"Whether the measurement is absolute or relative, the vast majority of kinetic studies of organic compounds measure the coefficient for the total reaction, based on the rate of consumption of at least one of the reactants or sometimes from the time-resolved analysis of products. While site-specific information – determined from the quantification of product(s) formed – is available in some cases, these data are not included here."

Issue: The referee finds the following statements confusing: "There are 800 rate coefficients in the current database where we have recommended extended Arrhenius parameters, based either on acceptance of recommendations from previous evaluations, or our own refitting of the data. For practical reasons, we did not re-fit data unless we judged that the existing recommendations did not encompass all measurements (which may include recently published data) or had other problems. This was done for ~50 rate coefficients, all of which were for OH reactions. For the remaining ~2000 reactions, recommendations were provided for rate coefficients around 298K, because of lack of data at other temperatures, or (in a very few cases) because the temperature dependence around 298K could not be described by the extended Arrhenius expression above."

Response and Action: We agree that these statements could be confusing and have rephrased them accordingly: "There are approximately 800 reactions in the current database for which the Arrhenius

equation has been used to describe the temperature dependence. In the majority of cases, temperature-dependent parameters were taken from previous recommendations, but temperature dependence was re-fitted where problems were identified, such as when temperature ranges were truncated to enable the simpler Arrhenius equation to be used, or where not all data had been incorporated into the recommendation. Data were re-fitted using the extended Arrhenius equation in ~50 cases, all of which were OH reactions. In total, there were 1951 rate coefficients for which only a room temperature rate coefficient was recommended. In the large majority of cases, this was because of an absence of data outside of room temperature. However, in some cases, such as where room temperature determinations were in agreement, but where temperature dependences were inconsistent, only the room temperature rate coefficient was recommended."

Issue: The referee asks whether the gas kinetic limit given on Line 159 is unique.

Response: We think that referee is questioning if the gas kinetic limit can be represented by a single value. If this is the case, then the answer is no, since it will depend upon the structure of the reactants, which is what we state in line 159-160. ("Reactions occurring on essentially every collision have a rate coefficient known as the gas kinetic limit, which is approximately 5.0×10^{-10} cm^3 $molec^{-1}$ s^{-1} at 298 K, although its precise value will vary with the structure of the reactants")

Action: This requires no further qualification and we propose no changes to the text.

Issue: The referee prefers that we state Arrhenius activation energy in line 193.

Response and Action: The sentence fragment "… the activation energy, E, was …" was changed to "… the measured Arrhenius temperature dependence parameter, B, was …". This refers back to Equation (2), which shows how A and B are used to obtain k(T). We agree that the use of "E" may be confusing because it was not employed previously (except to define "B").

Issue: Line 229: The referee would like to know if there is a reason for representing nitrogen–oxygen bonds as dative.

Response: There is no reason clear to us why the dative representation should be preferred. The point is to choose one and be consistent in this choice, such that SMILES strings can be represented uniquely.

Action: To make this clearer, the sentence fragment "where we prefer to represent these bonds as dative" is changed to "where we chose to always represent these bonds as dative, solely for consistency"

Issue: The statement in line 250 is confusing: " As shown in Table 1, of the 1644 compounds studied so far, most reactions have been measured for species that contain two or fewer functional groups

Response: The referee might be confused because the number "1644", the number of compounds with rate coefficient entries for any of the four oxidants, isn't actually on the table. Also, that number needs to be updated to 1564.

Action: To address this, the header on Table 1 is changed from "Numbers of compound types with rate coefficient entries" to "Numbers of compound types with rate coefficient entries (1564 compounds with data)" so the reader can see what the number is referring to. The reviewer may also be confused by our use of the term "functional groups", which could include hydrocarbon groups like methyl, ethyl, vinyl, phenyl, etc. The term "functional group" in this paper refers to *non-hydrocarbon* groups, such as -OH, carbonyl, halo, etc. To make this clearer, we change the sentence that first refers to Table 1 to indicate it refers to "non-hydrocarbon" functional groups.

Issue: PAN is not defined.

Response: This is correct. Since there are only two occurrences of this abbreviation in the paper, in this case we prefer to state the name of this compound type in full in both cases, that is, peroxy acyl nitrates, not polynuclear aromatic hydrocarbons as the referee suggested.

Action: The term "PAN" is replaced defined as "peroxy acetyl nitrate" wherever it appears in the text.

Issue: The referee finds the statement about free-energy relationships in line 336 is confusing: "… one well-established method of estimating rate coefficients that arises naturally from the compilation of data presented in our database is that of the free-energy relationships exhibited by rate coefficients of VOCs between different oxidants."

Response: The referee is probably confused about the use of "free-energy relationship" to describe log-log correlations of rate constants such as shown on Figure 7, because that term was not defined or used before in the manuscript.

Action: Since a discussion of free-energy relationships is beyond the scope of this paper, the term "free energy relationship(s)" was changed to "correlation(s)" in this paragraph and in the caption for Figure 7.

Reply to referee #2:

Specific comments

Issue: The referee would like to know if all data in the compilation have been evaluated, even where only a single measurement is available.

Response: Yes, even where a single measurement is available, we have done our best to check the plausibility of the values reported. This is not always easy to do, since this is not always clear for compounds with unusual functional groups or combinations of functional groups. Often, the reported measurement has been accepted, but the uncertainties have been increased to be more in keeping with the expectations of a given reaction class, for example, NO_3 reactions tend to have much larger overall uncertainties than the uncertainties that are reported (which often only account for the statistical uncertainties). However, because of the objective to compile as comprehensive a database as possible within the available amount of time, we necessarily had to spend less time evaluating each entry than has generally been the case for previous evaluations such as NASA and

IUPAC. That is why we consider this work complementary to those evaluations where important reactions are examined more closely, rather than superseding them, as discussed in the response below.

Action: The points above are included in our response to a related issue brought up by this reviewer, discussed below.

Issue: Have any reactions for which only a single measurement is available been excluded?

Response: Yes, in some unusual cases, a value has been reported that has raised concerns about the overall validity of the measurement. One example of this would be where an ozonolysis measurement has been conducted in the absence of (or with inadequate concentrations of) radical scavengers. This leads to unrealistically large rate coefficients that have no interpretable physical meaning.

Action: No changes made since this is already implied in our discussion.

Issue: The referee notes that checking values for 1000s of compounds could be very time-consuming and non-trivial. The referee further notes that given that our database encompasses the rate coefficient information contained within the IUPAC and JPL reviews, that it could be misconstrued as a successor and replacement for these efforts.

Response: The referee is correct, overall, that this was a big challenge for the project. However, many compounds in the dataset are quite straightforward, with the measured values being generally consistent with our expectations. A practical approach was taken, where a reviewer assessed a long list of rate coefficients and accepted, rejected or proposed changes to existing values in the unevaluated database. These actions were compared between reviewers, and where there was unanimous agreement, values were accepted into the database without further consideration. Subsequently, a shortlist of entries was made, where disagreements were encountered. These were then discussed on an individual basis until a resolution had been made.

The referee's second point about the potential of this database to supersede existing long-standing evaluation projects is also considered while addressing the first point.

Action: To address this and other issues brought up by this reviewer, we added the following paragraph before that starting on Line 191: "Since performing detailed evaluations for each reaction in a database of this size is time-consuming, a streamlined approach to the review process was taken, where a reviewer assessed a longlist of rate coefficients and accepted, rejected or proposed changes to existing values in the unevaluated database. These actions were compared between reviewers, and where there was unanimous agreement, values were accepted into the database without further consideration. Subsequently, a shortlist of entries was made, where disagreements were encountered. These were then discussed on an individual basis until a resolution had been reached. Although we consider this approach appropriate for our objective of compiling as comprehensive a compilation of evaluated data as possible within the available amount of time and resources, a number of individual reactions are discussed in more detail in the IUPAC, NASA, Calvert evaluations, and these are noted in our database as reviews where additional information can be obtained. We therefore consider this work to be complementary to these previous efforts, and

where detailed evaluations exist, readers are directed to the datasheets/ notes found within such publications."

Furthermore, we also added some clarification on this point on lines 52–54: "The JPL and IUPAC panels are a vital resource, providing detailed evaluations of the major inorganic and organic reactions of importance in atmospheric chemistry, and in fact their reviews of VOC oxidation rate coefficient data (although limited in scope) provide a starting point for our compilation. The work conducted here should be viewed as complementary to these activities, most closely aligned in scope with the Calvert et al. set of reviews."

Issue: E and R are not defined, when presented on line 123.

Action: we appended the definition of B as follows:

", where B = E/R, the ratio of the activation energy to the gas constant."

Issue: Lines 143-144: The reviewer makes the point that product formation can also be used to determine rate coefficients.

Response: We agree that total rate coefficients can be measured from the observation of the temporal profile of a product, and have re-worded the text accordingly (see also our response to Rev. #1 on this point).

Action: Text modified as follows: "Whether the measurement is absolute or relative, the vast majority of kinetic studies of organic compounds measure the coefficient for the total reaction, based on the rate of consumption of at least one of the reactants or sometimes from the time-resolved analysis of products. While site-specific information – determined from the quantification of product(s) formed – is available in some cases, these data are not included here."

Issue: The referee makes the point that there is likely to be a threshold in GECKO, where very minor reaction channels will be excluded, and that those compounds that are not contained in the mechanism for this reason are not truly "unknown unknowns", but rather "unexplored but potentially known unknowns".

Response: We think that referee is quite correct about the simplification of GECKO, and also has made a valid argument about the logic surrounding these three broad classifications of ours. There are some subtleties regarding these classes, one of which the referee has highlighted. There is at least one other possibility: "unknown knowns" which would represent any compounds for which there are kinetic measurements, and which are formed in a mechanism that is currently unrecognized. However, the caption for Figure 5 is quite large as it is, and it might not be appropriate to explain this kind of detail there.

Action: We therefore propose to enhance the discussion of these classes in the discussion starting on line 303, by adding the following paragraph after the sentence ending in "… are currently unknown to the GECKO-A model":

"Beyond these three main groupings, there are several other logical criteria by which species that are not contained within chemical mechanisms may be classified. For example, species that are formed through very minor reaction channels may be excluded by simplification protocols that aim to curb the combinatorial explosion within models, and may be considered as "unexplored but potentially known unknowns". Furthermore, for those species which have kinetic measurements, but have formation pathways that are currently unknown to chemical mechanisms, these may be considered as "unknown knowns". These groupings are, however, expected to be small in relation to the "unknown unknowns.""

Issue: The referee states correctly that alpha-pinene is not addressed directly in Aumont et al. (2005), and makes a good point that for a comparatively simple terminal C7 alkene, almost as many isomers are produced after 5% have been trimmed through simplification.

Response: The purpose of citing Aumont et al. (2005) in this case was to reference the GECKO software in general, rather than a specific compound or indeed a specific version of the software. However, there is a more appropriate citation to use in this case: Valorso et al. (2011), who investigated alpha-pinene directly. The amount of species generated in by GECKO after the mechanism reduction protocol has been applied is ~400,000 (based on the favoured Nannoolal–Nannoolal vapour pressure estimation method), which matches the value that we have reported in our paper. Nevertheless, our explanation of this number was inadequate, partly because of our citation choice and partly because we did not explain that this value was itself the product of a simplified mechanism.

Action: The sentence on page 2 now reads:

"For example, in the GECKO-A model, the number of possible products formed from a single VOC of only intermediate complexity, α-pinene, can result in ~400,000 different species (Valorso et al., 2011), after a mechanism reduction protocol was applied."

Reply to referee #3:

Issue 1. " The database does not contain all types of atmospheric reactions of organic compounds"

Response: Although what the referee says is true, when we were considering titles for this article, we concluded that this database may not be restricted to these four oxidants throughout its lifespan, and that at least some other oxidation processes may be included in subsequent iterations. This would render a more descriptive title such as the one proposed by the referee obsolete and may lead to subsequent confusion regarding the limitations of this resource.

Action: For the reasons outlined above, we would like to retain our current choice of title.

Issue 2. Rephrase sentence thus: Organic compounds with a large variety of functional groups are present in this database.

Action: Sentence changed to: "Many compound types are present in this database, including naturally occurring chemicals emitted to or formed in the atmosphere, and anthropogenic compounds such as halocarbons and their degradation products."

Issues 3-14: Various minor editorial suggestions and corrections

Response and actions: We agree with the referee in most of these cases, and corrected the text accordingly. This includes minor corrections to the instruction manual and spreadsheet noted in issue #14. We thank the reviewer for his/her careful proofreading. The only exception is that we did not change the word "troposphere" to "atmosphere", because our database is primarily relevant to the conditions of the *lower* atmosphere.

Issue 15. "It would be great if users could contact the database managers to give suggestions etc."

Response: This is a good suggestion, and we have been considering something along these lines. Our current thinking is to set up a mailing list that users can subscribe to on the website, and this way they can be alerted to updates as well as being able to provide feedback. We wanted to gauge the level of interest before committing to a single course of action.

Action: In the meantime, a "Contact us" section will be added to the Eurochamp web page where the data are being made available, giving an email address people can use to communicate with us about the database. There are already links giving email addresses of the contributors, but this will make it clearer that we welcome such contact, and provide a central contact point, which initially will be Max McGillen, the corresponding author.

Issue 16. "… the database will need a name …"

Response: Naming the database has caused our group some mild consternation. Unsurprisingly perhaps, it would appear that attempting to name something by committee decision is not a very productive activity. However, we do not believe the name to be of essential importance to the project at present.

Action: No changes made.

Reply to referee #4:

General comment: One general comment regarding the database behind the article is related to the rate coefficients presentation and the link with the initial article which published the kinetic results. The database is fully functional and it is a perfect collection for the researchers working on SAR methods and chemical models. However, the database is constructed on the results provided by many articles which deserve as well the citations as the database itself. I suggest to link in a more visible manner the references which are providing the rate coefficients values for the database. Please find a way to add in Excel spreadsheet for the recommended kinetic parameters a hidden comment easily visible by selecting the value including the reference articles which are linked with that rate coefficient value.

Response: we strongly agree with sentiments of the reviewer that the data contained within this database are the product of many researchers who deserve recognition for their time and effort. However, we would like to point out that the current version of the database makes it quite simple for users to extract the bibliographic information that would be required to cite individual studies/

reviews. Presently, this doesn't apply to the multiple studies that are considered within existing reviews, but this could represent a future improvement for the database if this becomes important for our userbase.

The way in which to locate citation information for the papers that report measurements is by selecting the "Find Compounds" worksheet, locating the compound of interest using one of the search criteria, select the row of interest (if more than one compound is listed), and clicking on the "info on compound in selected row" button. Selecting the "Selected Compound Info" worksheet provides all the available bibliographic information.

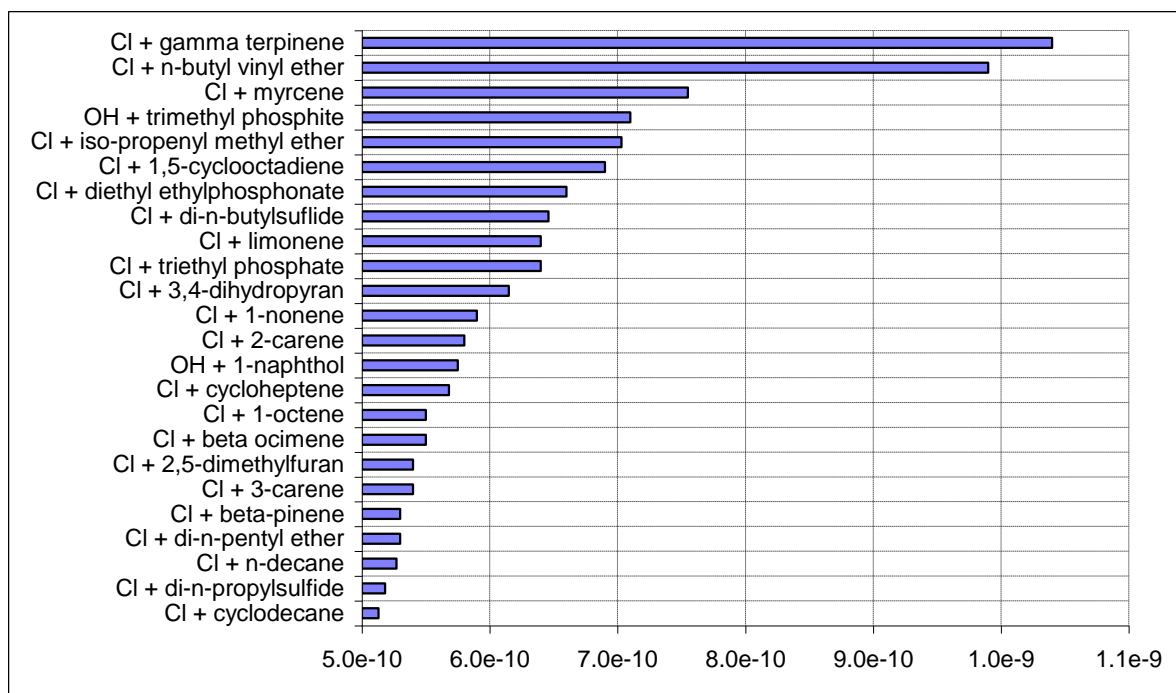
Action: No changes made to the manuscript itself. The following short paragraph is added to the end of the section in the instruction manual discussing how to listings or recommended kinetic parameters: "The rate parameter listings output on the "Compound List Info" sheet do not give the citations or other information we have in the database for the compounds and reactions shown. This information can be output using the "Selected Compound Info" macros and worksheet, discussed below."

Issue: The referee disagrees that relative rate studies must use a rate coefficient that has been determined by an absolute technique. (Line 99-100)

Response: We generally agree with the referee on this issue. However, we would like to point out that in the scenario where a reference rate coefficient had been determined against another relative rate, which has itself been determined against a further relative rate, etc., the propagated uncertainties would increase rapidly in each step of this process, until they become unacceptable to our review.

Action: The sentence in question was changed to read: "In relative rate studies, the rate of the reaction of interest is measured relative to that of a reference reaction whose rate coefficient has been placed on the absolute scale."

Issue: The referee has some concerns about the example that was used for the gas kinetic limit, and points out that there are larger rate coefficients contained within the database than the long-chain hydrocarbons. (lines 159-164). The referee also questions our acceptance of the high rate constant for Cl + 1-etheneoxybutane.

Response: We consider a detailed discussion of the gas kinetic limit to be beyond the scope of this paper, but agree that maybe our presentation of examples could be improved. The reason we include the high rate constant for Cl + 1-etheneoxybutane and other compounds is that they were measured experimentally, and we believe we should reject such data only if there is experimental evidence to do so. As stated in the manuscript, the value of the gas kinetic limit "will vary with the structure of the reactant". The limits themselves are determined by calculated reaction cross sections which depend on several non-trivial assumptions about the geometry and radii of the reactants. So these are not hard-and-fast limits. However, because of the concern this reviewer had about high rate constants in our database, we reviewed the highest recommended rate constants on our database, and found 24 rate constants above 5×10^{-10} cm^3 $molecule^{-1}$ s^{-1}. These are shown on the graph below:

The results show a number of measured rate constants between 5 and 8×10^{-10} cm^3 molecule^{-1} s^{-1}, for a variety of compounds from several laboratories, using different methods. Therefore, it is unreasonable to reject rate constants in this range, and it appears that even 8×10^{-10} cm^3 molecule^{-1} s^{-1} shouldn't be considered necessarily incorrect. However, the Cl + 1-etheneoxybutane and gamma-terpinene might be outliers. For this reason, we revised our database to indicate that these rate constants are upper limits. The information will still be available to the users, but will not be included by default when multiple rate constants are requested.

Action: (1) The paragraph has been modified and some discussion of very large rate coefficients is now included "Reactions occurring on essentially every collision have a rate coefficient known as the gas kinetic limit, which is approximately 5.0×10^{-10} cm^3 molecule^{-1} s^{-1} at 298 K, although its precise value will vary with the structure of the reactants. The recommended rate coefficients at 298 K in the present compilation span the range from the gas kinetic limit for the reactions of chlorine atoms with several species, to less than approximately 10^{-22} cm^3 molecule^{-1} s^{-1} for the reaction of ozone with halogenated alkenes which corresponds to reaction in approximately 1 out of 10^{11}–10^{12} collisions. We find that several laboratories have reported rate coefficients (mostly involving reactions with atomic chlorine) that are considerably larger than would be expected from a simple collision theory calculation. For reactants with large dipole moments such as Criegee intermediates, rate coefficients in excess of the collision limit have been rationalized (Chhantyal-Pun et al., 2017, 2018). However, it is more difficult to explain such high rate coefficients in the chlorine reactions, and it is possible that further measurements and theoretical work may be helpful in this regard." (2) The 298 K rate coefficients given for Cl + gamma terpinene and Cl + n-butyl vinyl ether (1-ethoxybutane) are flagged as being upper limit values. People searching the database will find that there is information about these rate coefficients, but they have to use the use the "get info on selected compound" macro to see this information, provided with a comment about the uncertainty.

Issue: (Line 104) please consider other techniques as well (e.g. PTR, monitors) and not exclusively GC and FTIR.

Response and Action: The referee is quite correct that there are many more analytical techniques that are available to kineticists, and we have rephrased this sentence as follows:

"The reactant and reference compounds are monitored using one or more of the many chromatographic, spectroscopic and mass spectrometric techniques that are available. "

Issue: (Line 104) Please mention that should be there a "near zero intercept" for the slope.

Response and Action: Yes, this is always preferred. We have amended the sentence as follows:

"A plot of $\ln([VOC]t_0/[VOC]t)$ versus $\ln([reference]t_0/[reference]t)$ has a slope equal to the rate coefficient ratio $k_{VOC}/k_{reference}$, where t_0 and t refer to initial concentrations and concentrations at time t respectively. This plot should be linear and intercept the origin, indicating that secondary chemistry is not significantly affecting the concentrations of the VOC or the reference compound."

Issue: The referee prefers "standard Arrhenius equation" as opposed to "Arrhenius equation".

Response: We understand that the referee is trying to further distinguish this expression from the "extended Arrhenius equation", however, the term "Arrhenius equation" is so widespread and well-understood that it would seem unnecessary to inject "standard" into this commonplace term.

Action: No changes made.

Issue: (Eq 2) Please describe E for the equation since you described A (pre-exponential factor).

Response: This was an error already brought up by Referee #2, which we have addressed.

Issue: (Line 331) I suggest replace "daughter" with "secondary"

Response: Upon reflection, we would prefer to replace "daughter" or "secondary" with "oxidation". The trouble with "secondary product" is that it could be misinterpreted as a product formed from two oxidation steps. In fact, there could be a far greater number of oxidation steps occurring for some compounds, and it therefore seems more appropriate to use "oxidation product" instead.

Action: " … will generate an immense number of daughter products under atmospheric conditions" changed to " … will generate an immense number of oxidation products under atmospheric conditions."