Response to reviewer 2

Many thanks for your extensive and valuable suggestions. We have amended the text in response to your comments, as detailed below. Note that this response in some cases overlaps with the response to reviewer one, so we have not distinguished text changes to this review from those for the other review. The format is as follows:

Original comment

- Any responses
    - Description of changes

Changes made

1a. In most parts of the manuscript and in the source code the specified aim is to complete missing emissions in future scenarios produced by IAMs. However, in some sections of the manuscript, other aims are indicated (e.g. complete stylised scenarios, fill missing sectoral data, use historical estimates, aggregate regional data(?)) which I found confusing sometimes. One solution could be to remove these hints of other applications and use the (IAM/stylized) future scenarios as the aim and example throughout the paper. Additional applications could then be outlined in a discussions section (in a bit more detail).

    - To reduce the confusion, we have removed the reference to aggregating regional data entirely and have rephrased the section around historic estimates and sector emissions. Since one of the crunchers (last time ratio) is specifically designed to deal with the case of historic emissions, it is still discussed there. We mention regions in the context of structure as follows: “Pyam dataframes assign values to variables as a function of different models, scenarios, regions and times. All methods work on databases with only a single region at a time, although the region can be different between the infiller and target databases.”

1b. With respect to the main aim I would have expected use cases showing the completion of several variables of different IAM scenarios.

- As mentioned in the separate discussion piece submitted previously, we are somewhat confused by this comment – the first “use case” provides a case where a scenario (now a different scenario) has many variables completed, the second and third cases complete variables from multiple scenarios. There is also a github repository consisting only of examples of use (some of which are mentioned in this paper, some not – note that the Sky scenario example in the text has changed now) and a “notebook” section to the main github talking through the general principles of applying these techniques. Could you clarify if there is something else you would like to see added, and whether you really want this included in the paper itself? (We are essentially treating the examples github as a living and evolving Supplementary Info, although the archived version is always available from Zenodo.)

2. I would recommend to rework parts of the abstract and of the introduction in order to better cover the content/ set the scene for the rest of the manuscript (e.g. add info about rank correlation, tests/ use cases).

    - We now mention the use-cases in the abstract and have an outline of the manuscript in the introduction. “A variety of infilling options are outlined and their suitability for different cases are discussed. We recommend certain infilling

techniques as the good defaults, but emphasise that considering the specifics of the model being infilled will produce better results. We demonstrate the package's utility with three examples: infilling all required gases for a pathway with data for only one emission species, splitting up a Kyoto emissions total into separate gases and complementing a set of idealised emissions curves to provide a complete, consistent emissions portfolio."

3. Literature: it is rather difficult for me to imagine that there are no other somewhat comparable tools around and that so far missing emissions were usually set to zero or only somehow unsystematically filled following 'ad hoc' decisions (as stated in the introduction). For completeness it would be nice if the authors could dig some more into the literature and check how climate models so far got the required input from IAMs? One example for a tool covering a similar purpose in maybe a slightly different but connected setting is the tool used in Gütschow et al. (2016) and du Pont et al. (2016) which is described in Nabel et al. (2011). Some of the co-authors have been involved in these papers.

- As discussed with Robert Gieseke in previous correspondance, some specific papers you mention do not strictly do infilling – they use non-emissions economic data to perform the analysis, whereas we deem 'infilling' to be when no such data is available.
    - We have changed the tense of the sentence to clarify that explicitly setting these values to zero is rare, although very often done implicitly by ignoring the variable altogether. In explicit cases, the Equal Quantile Walk is used. We now mention Nabel et al as an example of using this but note that there is no general tool for this so we think our open-source codebase provides a significant step forward in terms of reducing duplicated effort. We also now highlight the use of a RMS-closest technique by Robiou du Pont & Meinshausen, 2018. The text now reads:
    - "Most earlier studies overcame this problem in one of two ways: with expert-based ad-hoc decisions on how to adequately fill-in missing species (Schaeffer et al., 2015); or by assuming that a pathway will occur at the same quantile for each set of emissions in a particular year, although the quantile can vary over time (Gütschow et al., 2018; Meinshausen et al., 2006; Nabel et al., 2011). However, the former clearly does not scale easily to larger databases (because making ad-hoc decisions for a thousand scenarios requires a significant time input), and the latter approach, termed the "Equal quantile walk" (EQW), ignores trade-offs and specific relationships between emission species resulting from how competing technologies, behaviours and industrial practices result in different emissions. A few alternative approaches have been used recently: for instance, using the pathway with the smallest mean-squared distance over all time was used in (Robiou du Pont and Meinshausen, 2018). This works well for large databases containing similar paths, but is less reliable for smaller databases or for paths with an unusual behaviour over time. A more sophisticated "Generalized Quantile Walk" technique can capture the effect of trade-offs and was recently introduced in section 3.8.1 in (Teske et al., 2019), involving quantile regression between a lead variable (fossil CO_2 emissions) and other gases for every individual year. Unfortunately, the implementation there did not consistently guarantee that higher quantiles resulted in higher emissions, and has not been followed up with any peer-reviewed work that does so. A tool for infilling was provided with (Rogelj et al., 2014) using a cubic spline between specific points in a small database, however this type of infiller behaves chaotically when

applied to large databases incorporating many different models. It was also coded in Excel, limiting the ease of open-source development."

4. While reading I sometimes got confused by different terms and I found parts of the manuscript a bit sketchy or difficult to read. More specific:
a) There are several changes in terminology among different (sub-) sections (e.g. "lead variable", "lead gas" "inputs or outputs" (l.89), and timeseries; and a sudden switch to model and scenario in 2.2.2 where 2.2.1 only had the more general term database; but also small things as the change from CH4 in section 1 and 2 to methane starting section 3). I think for the reader it would be helpful to stick to a certain terminology throughout the manuscript.

- References to "lead/follower gas" have been changed to "lead/follower variable". Inputs is now generally removed where it is equivalent to "infiller", see below. Timeseries refers to any data that changes over time and is used to refer to any of the above – it is a data structure, not a description of the data meaning.
    - A short paragraph explaining what model/scenario combinations are has been added before 2.2.1: "As one final detail, we discuss the data model which is assumed by Silicone. Silicone is built around the pyam package (Gidden and Huppmann, 2019). As a result, it assumes that all input data is provided in a particular structure. The structure includes the model which created the timeseries, the scenario with which the timeseries is associated (e.g. a high BECS 1.5 degree scenario), the region the emissions occurs in and the unit of the data (full details available at https://pyam-iamc.readthedocs.io/en/stable/data.html). Accordingly, Silicone is able to work on specific subsets of models (e.g. only the MESSAGE model) or subsets of scenarios (e.g. all SSP1-like scenarios). We therefore follow the pyam convention and refer to a "model/scenario combination" to mean a single projected world, that in some contexts might be called a "scenario"."

b) The terms infiller and infilly are very difficult to distinguish in quick reading and I think it would help a lot when choosing less similar terms – How about source and sink/target, or infiller and target, or infiller and silicon-filled, or comparable.

- Infiller and target seems like a good combination, now used throughout.

c) In subsection 2.2.2, a bit out of the sudden, several scenarios and models seem to be presupposed, while at the beginning of section 2 only "a database that contains data for at least two emission species" is kind of officially introduced. Maybe it would help to directly introduce the use of different models/IAMs and their scenarios at the beginning of section 2 such that the usage of different timeseries (2.2) and different models and scenarios (2.2.2) is less out of the sudden. An alternative could be a consistent use of the more general terms of "timeseries of different lead variables in the infiller database" depending on the main aim of the paper (see also point 1).

- Change made as described in 4a), introducing the structure of pyam dataframes.

d) Please consider to better structure 2.2.1 and 2.2.2 (e.g. print algorithm names in bold or in italics, with separated paragraphs for the different algorithms or /and as lists (e.g. latex 'description' or the like)).

- Sub-subtitles now used, in bold.

e) Equations are throughout embedded in the text (see also point 5).

- Key equations now in "display" mode.

5. Since the different algorithms for the completion of emission timeseries are the main scientific contribution of the paper I would appreciate if the equations could be clearly separated from the text (i.e. introduced as separated numbered equations) and, furthermore, if more equations would be added (see also specific comments below). In my opinion this could increase readability (4e) and reproducibility. Ideally it could also help to better understand how cases of several lead variable pathways are treated in the different algorithms.

- As above. Important equations are numbered for external reference and we have added several equations for clarity.

6. Test and use cases only show the usage of absolute value based algorithms, which I find unfortunate and a bit incomplete given the share of the method section dedicated to the ratio based algorithms. How about at least including examples using the time dependent ratio method?
    - One of the ways of splitting the Kyoto gas totals is the "decompose collection with time-dependent ratio", which is a wrapper calling the ratio-based method several times. We have modified the text to make this clearer that this such an example.
7. Please check the format of your references in the text (e.g. l.28, l.59, l.60, l.68, : : :)
    - This has been corrected.

Specific comments/questions

l.1 Why Silicone?

- Our package fills what some would call gaps in emissions scenarios. Silicone is a caulking agent used to fill in gaps in tiling and was the first thing one of the developers thought of when we searched for a name. We are happy to add an explanation in the text if you feel it is necessary.

l.16 Transition. E.g. "In this paper: : :"

- "This paper presents a variety of infilling options and outlines their suitability for different cases are discussed. We recommend certain infilling techniques as the good defaults, but emphasise that considering the specifics of the model being infilled will produce better results. We demonstrate the package's utility with three examples: infilling all required gases for a pathway with data for only one emission species, splitting up a Kyoto emissions total into separate gases and complementing a set of idealised emissions curves to provide a complete, consistent emissions portfolio."

l.16 Please consider to add more information here about the content of the paper

- Detail added, as above

l.33 : : : exert : : : between? Please check language

- Changed to "as a large number of supposedly minor emissions may collectively exert a significant radiative forcing."

l.36 Is there an example reference/study where filling with zeros has been applied?

- Implicitly, this is done by every study where any of the F-gases are ignored (e.g. every IAM in the SR1.5 database ignores NF_3). However, studies don't tend to list the gases they ignore.
    - "If no infilling is attempted, the unevaluated emissions would effectively be considered zero, which would clearly create systematic biases and potential artefacts in the projected temperatures."

l.41 I do not understand "does not scale easily"

- Added "However, the former clearly does not scale easily to larger databases (because making ad-hoc decisions for a thousand scenarios requires a significant time input)"

l.45 There is no 3.8.1 in this reference (reference currently points to Chapter 4, if you target 3.8.1 consider citing Teske et al.)

- Corrected as advised

l.60 Please specify what "this" refers to

- Now "Silicone"

l.64 "suite of tools" are these all python tools?

- Indeed, "Python" added

l.66 Are there more than these two? Else consider to add "and/as well as" before "harmonise..."

- The pipeline can be considered to include the OpenSCM part, although this isn't strictly developed by the IAMC, it's therefore a little fuzzy.
    - "This pipeline includes tools to manipulate and plot IAM data (pyam, (Gidden & Huppmann, 2019)) and harmonise mismatches in historical emissions (aneris, (Gidden, Fujimori, et al., 2018)). The estimation of climatic impact is performed by OpenSCM, managed by the OpenSCM community (Nicholls, Gieseke, Lewis, & Willner, 2020), which is compatible with the data structure of the pipeline."

l.67 Consider deleting ", managed by the OpenSCM community"

- As above, we have revised the sentence slightly according to this comment, while the text still makes clear that this is not part of the IAMC

l.82 several? Three/two: : :

- Currently 4 (one due to an update from the original paper draft), more may well emerge. One is arguably an "aggregation tool" rather than a "multiple infiller". We maintain ambiguity for this reason

l.85 Maybe change to "Currently, there are ..."?

- "Currently" added

l.98 Consider rephrasing e.g. "... and where emissions are expected to scale with each other ..."

- Rephrased, this discussion is moved to the next paragraph. "The ratio-based approaches are better for cases where the lead values to be infilled are outside the range in the infiller database and we expect the emissions to scale with each other."

l.98 What do you mean with "regional data" and "aggregate data"? If you refer to regions as subset of global data then this is the first time that a spatial reference is given and I wonder if it would be appropriate to introduce this more formally earlier in the manuscript?

- Reference to regional data has been removed as discussed above. A reference to "splitting up an aggregate basket of emissions" mentioned above, this case now changed to "splitting up aggregated emissions into their components"

l.100 Consider to give an example? CO2 uptake?

- Added ", e.g. CO_2 emissions"

l.106 What do you mean with similar – similar magnitudes?

- Indeed, clarified to ", preferably with both larger and smaller lead emissions in the infiller database"

l.111 "estimate the ratios" – if not to be predefined: : :

- Now "determine"

l.112 "follower value in infillee database" -> "in the ..."

- Added as requested

l.112 Please consider to visually separate (and number) the equations.

- Change made for all significant equations

l.114 "each different timeseries" -> different regarding what - do you mean different follower variables?

- Sentence removed.

l.116 mean regarding what - time or different sources (models, scenarios?)

- Sources, clarified to "The 'latest time ratio' method uses the ratio between the mean follower data in the infiller database (we denote this database with lower-case, e_f) and the value of the lead variable in the target data (E_l)"

l.117 what does "both" refer to?

- Both means follow and target lead. Now "both values"

l.119 what do you mean with "all estimates"? Different sources (models, scenarios)?

- Yes, specification of "at that time" added

l.120 why historical? Couldn't this also be different scenarios from different IAMs?

- It can, it's simply that this is the most common use-case (we are yet to see a case where all IAMs model a gas up to a given point then all stop thereafter whilst continuing to model other gases).

l.120 "and the lower case ef (t) represents the follower values in the database at time t." -> "and the lower case $e_f(t_{last})$ represents the follower values in the database."?

- Rephrased as requested, the distinction being between *E* and *e.*

l.125 the infillee lead is not part of the formula – is this the El(t) from l.113?

- Yes, it will be used to multiply the ratio.

l. 125-133: Maybe consider to restructure? You could start with the context, i.e. the algorithm name, and then add the explanation, e.g.: "The decompose collection multiple infiller is based on: : : relying on the useful property of : : :"

- Significantly restructured and added more equations as explanation

l.134-137: Equation for R(t)?

- We explain that this follows the logic below instead, it would be tedious to write out here again with one symbol different. The whole passage has been rewritten.

l.142 I did not understand this sentence

- Rephrased to "This is the only infiller that is not time-independent, i.e. changing the value of the lead at one time may result in different outputs at other times."

l.143-145: Equation for Ef?

- We have added it.

l.151 There is no 3.8.1 in that reference (reference currently points to Chapter 4, if you target 3.8.1 consider citing Teske et al.)

- Well spotted, resolved as above

l.152 There seems to be a lost copy of the figure caption in the text. (Either just delete or maybe rephrase to steps with complete sentences)

- Indeed, deleted

l.155-165 Equation for Ef?

- This would be illegibly long and complex without a great deal of separately defined objects. The whole section has been rewritten however.

l.163 Maybe Ef not El?

- The weights are associated with e_f, the ordering is associated with e_l. This has been rewritten, as noted above.

l.166 What about the KyotoGHGs as one basket?

- Could be considered (and can be calculated by the code in the examples github repository if you're interested), but interpreting it is complicated by the fact that
- 0) calculating the Kyoto GHGs as a single basket only makes sense if all relevant Kyoto gases are reported, which isn't often the case (if it were Silicone's domain of applicability would be severely reduced)
- 1) there are several possible metrics for this, on top of which the values stored in the IIASA database for a given metric do not entirely equal the values calculated from their components. Sometimes this is due to the incompleteness issue, other times there appear to have been rounding errors or method disputes in the process.
- 2) it naturally correlates very well with its primary components CO2 and methane without really signifying anything, further complicating the averages analysis.

l.169 Maybe give an example for two such variables?

- Added "for instance black carbon and carbon monoxide are both produced by incomplete combustion"

l.185 Consider explicitly listing the two constituents

- OK, it is somewhat long-winded to spell out the acronyms here, so we simply use them and hope those unfamiliar with them will look at the table

- "(AFOLU and Energy and Industrial processes, a similar concern can be raised about F-gases)"

l.186 Which are "these two" – CO2 and CH4?!

- Yes, spelt out now. "CO_2 and CH_4"

l.188 you write that BC, CO and OC "correlate poorly with others, however, from the table it seems that they do not correlate less well with others than other gasses, the main difference is that they correlate very well with each other.

- Clarified: "that correlate well with each other but less well with other emission pathways"

l.190 maybe aggregate F-gas emissions / F-gasses as a basket?

- As a basket added to the analysis

l.191 up to here always "CH4"

- Changed to always be CH_4

l.198 consider deleting "and find similar results"

- Ok

l.199 "we choose four" – which are basically all?

- Following from the point below, this now reads: "We use the crunchers that are designed for use on complete datasets with only default settings: QRW (default settings mean in absolute mode and for the 0.5 quantile), RMS closest, EQW, time-dependent ratio and linear interpolation. Interpolate selected model behaves identically to linear interpolation with default settings and is not treated separately here."

l.199 But even if there are errors, wouldn't it be interesting to see what happens? I would appreciate if you could also show results for the 'time dependent ratio' algorithm

- We now show the results from using the time-dependent ratio too, and slightly change our normalisation to make the comparison with "just use the mean" clearer. In several cases it does indeed give values worse than 1 (i.e. using the mean). A complicating consequence of this new normalisation and higher possible values is a strong skewing of the results and non-normal distribution of the errors. We therefore also substitute the Wilcoxon t-test for the student t-test in all cases. The statistics are all robust to this and it has no impact on our conclusions, although all of the p-values are slightly different now. References to Time-dependent ratio have also been added to the conclusion here.

l.201-202 Equation?

- The (new) equation has been added: "i.e. $\langle\sqrt{\langle\left(\frac{E_{f,inf}-E_{f,act}}{\sigma}\right)^2\rangle_i}\rangle_{decade}$, with the subscript text inf indicating that the value is infilled, act indicating actual and $i/decade$ indicating averaging over model/scenario cases or decades."

l.203 "both cases" – CO2 and CH4?

- Indeed, clarified

l.203 "non-CO2 pathways" – but CO2 is derived with CH4 -> maybe replace by "emission pathways"

- We have replaced this with "follower pathways"

l.204 I would rephrase this, if QRW would be fairly similar all four would be?

- Rephrased as "the next smallest"

l.208 What do you conclude from the non-Gaussian distribution test?

- Originally we concluded that we could use the Student's t-test. With the new normalisation and data, we now fail this test so use the Wilcoxon t-test. The qualitative conclusions are as before.

l.212 "either of CO2 or CH4"

- Corrected to "either CO_2 or CH_4"

l.214 capital T for Table 3

- Corrected

l.225 Add Silicon -> "Data in the Silicon package“

- Clarified to “in the Silicone examples package”, the main Silicone package is not dependent on any fixed data source. (Other data sources are used by the examples package, but are also included in it.)

l.238 treatment of regions has not been introduced, maybe explain better or consider deletion?

- Regions are now mentioned above, but not here, as described above.

l.243 “this database” – which? The SR1.5 repository?

- This whole section has changed, the data here now comes from a different (named) database, AR5.

l.249-252: Again a lost copy of a figure caption

- Deleted

l.285-287 “free variables” are mentioned twice but are not further explained?

- They are not tremendously relevant to the discussion here, but this has been changed to “in this case, based on rates of transition between the RCP pathways and a long-term emissions value”

l.287-288 I do not understand this sentence

- Changed to “Silicone provides an alternative means of complementing such results – instead of specifying the functional forms of all emissions, you can have a few key emissions prescribed and infill the remainder using scenarios with similarities to the desired narrative.”

l.300 consider deleting “of which there are many” or maybe replace by “several options” or the like

- Removed

Table2: Please explain the asterisk again in this figure caption

- Added “Names followed by asterisks use a ratio-based approach, i.e. they find a multiplicative factor and then multiply the target lead by this value, if the asterisk is in brackets there are ratio-based options.”

Table3: Consider to change the colouring – to me the yellow/orange highlighting gives a ‘positive’ impression. Maybe you could colour the cells with bold numbers in green and those which are currently yellow in red?

- The table has been recoloured. We have avoided using green and red for colourblindness reasons, but yellow for strong correlation and blue for high time-variability should perform the same roles.