We thank the reviewers for their comments and suggestions. Compiled below are responses to both reviewers' comments (in italics).

Reviewer 1

"Trying to reproduce experiment "exp_d1" (former terminology) I noticed one difference concerning the initial value for parameter "detmartin" between the provided configuration file "OxfordMOPS_exp_d1.json" (where the value is 0.5) and the provided experimental output in folder "OxfordMOPS_EXP/exp_d1" (where it seams to have been 0.54)."

We thank the reviewer for this comment. Yes this is correct, the file "OxfordMOPS_exp_d1.json" does request DFO-LS to initiate from a detmartin value of 0.5. However, as seen in the Supplement/OxfordMOPS_EXP/exp_d1 files (prog_exp_d1.txt and dfo001/parameters_input.txt), DFO-LS actually initiates from 0.54. This is because DFO-LS cannot start too close to a parameter bound (determined by the DFO-LS setting "rhobeg"), and therefore shifts the starting values away from boundaries if necessary. DFO-LS always prints a warning when it does this, so the user is always notified, and OptClim ensures the parameter input file (e.g. dfo001/parameters_input.txt) contains the shifted parameter values.

This is fully reproducible. If OptClim were to re-run DFO-LS using "OxfordMOPS_exp_d1.json" exactly as it is, DFO-LS will always shift detmartin from 0.5 to 0.54, and the results will always be as in Supplement/OxfordMOPS_EXP/exp_d1.

Reviewer 2

1. *"I should have been clearer in my first review point 38:*
   Eq. (30) + many lines: "fglobal". Usually non-variable subscripts are typeset upright. This comment should have mentioned all non-variable subscript and superscript. E.g., I would replace f^{Base}_\textrm{T} and r^{Base}_{qj} with f^\textrm{base}_\textrm{T} and r^\textrm{base}_{qj} (i.e., fBase and rBase with fbase and rbase) where I also avoided having a capital "B" for consistency. The authors should check the entire manuscript for any non-variable subscript/superscript and correct them."

   We thank the reviewer for this comment. We have done as suggested.

2. *"As per GMD's guidelines (https://www.geoscientific-model-development.net/submission.html#math) vectors such as x, which I guess was LaTeX'd from \mathbf{x} (first appearance l. 117) should be typeset "in boldface italics", i.e., x. This is easily done using the \vec{x} command provided by the Copernicus LaTeX template."*

   We thank the reviewer for this comment. We have done as suggested.

3. *"Units are missing in almost every figure and should be added."*

We thank the reviewer for this comment. Units were missing from 4/13 of the figures (Figs 4, 9, 11, 13), but have now been corrected. The misfit shown in other figures is unit less.

4. *"l. 178: Eq. (2): the sums should not start at i=1. They should be written as \sum_{i\in j} instead of \sum_{i=1}^{i\in j} (Σi∈j instead of Σi=1i∈j)."*

   We thank the reviewer for this comment. We have done as suggested.

5. *"l. 132: Mathematical symbol D should be in italics, i.e., D. (Use "D".)"*

   We thank the reviewer for this comment. We have done as suggested in Equation 1 and Line 123.

6. *"l. 168: "the length of the vector r" should be spelled out for clarity with maybe something like "d, the number of ri terms". (Also note that otherwise the vector r, which should be boldface italic, is not even defined.)"*

   We thank the reviewer for this comment. Line 154 has been corrected as suggested, from

   "There is no maximum suggested length of the vector ***r***, therefore …"

   to

   "There is no maximum suggested value for *d*, the number of r_i terms, therefore …"

7. *"Table 1: In retrospect, my suggestions for experiment names were not great. For readability, I think it might be better to have shorter names and avoid underscores. What about:*
   "CTL" for the CMA-ES run (the control run),
   "$SMOOTH_1$" for the "D_smooth_1 run" and so on,
   "$NOISY_1$" for the "D_noise_rand1" run and so on, and
   "$SPARSE_1$" for the "D_smooth_sparse_1" run and so on?"

   We thank the reviewer for this comment. Yes, we also agree that determining the best naming convention is difficult, as there is a balance to be found between containing the necessary information to describe each experiment, and keeping them short enough for good readability. It might be best to keep in the information as to which experiment is using which optimiser. Therefore, possibly a mix of the reviewer's previous suggested naming convention and the most recent suggestion above:

   "C_SMOOTH" for the CMA-ES run (the control run),

"D_SMOOTH$_1$" for the "D_smooth_1 run" and so on,
"D_NOISY$_1$" for the "D_noise_rand1" run and so on, and
"D_SPARSE$_1$" for the "D_smooth_sparse_1" run and so on.

We realise this doesn't fully eliminate the underscores, but it is still an improvement to the 3 underscores of D_smooth_sparse_1 etc., and allows us to include the optimisation algorithm information. We have revised to now use this naming convention, and we hope the reviewer deems them suitable.

8. *"l. 221: Use the \times symbol rather than the letter "x"."*

   We thank the reviewer for this comment. We have done as suggested.

9. We thank the reviewer for the comments below.

   9A. "Fig. 4:
   Maybe a line for the target value could be added in the background? (and a +-5% band?) "

   Thank you, this was left out in error and has been fixed.

   9B. "Maybe show the 10 CTL (C_smooth) starting points as tiny dots?"

   Thank you, this has been done as suggested, and the caption updated to explain this.

   9C. "There is a lot of unused vertical space in each panel. Maybe the y-axis limits can be tightened a bit? E.g., Fig. 4c shows maximum K_PHY values of about 0.2, but the y-axis goes up to 0.5."

   Thank you. The y-axes are fixed to the parameter bounds, within which DFO-LS was allowed to search, and allows the reader to understand where the starting and optimised parameter values fell within the bounded parameter space. Also, by adding the starting CMA-ES locations as suggested by the reviewer above, there is now less unused vertical space in each panel. For these two reasons the y-axes limits have been left unchanged, and we hope the reviewer agrees with this decision.

   9D. "The legend could be simplified to only say that circles are starting points and crosses are optimized values? (Maybe use black for the legend and then give a different color than black for the smooth values."

   Thank you, this has been done as suggested.

   9E. "Speaking of color, a color-blind-friendly palette could be used here instead of plain black, red, and blue (e.g., colorbrewer's qualitative colors (https://colorbrewer2.org/#type=qualitative&scheme=Dark2&n=3), but there are many others!)"

Thank you, this has been done as suggested in Figures 4 and 5. The colours in all other figures remain unchanged because the various marker styles ensure they are colour-blind friendly.

9F. "The legend could be placed at the bottom rather than in the middle to avoid visually breaking the x-axis alignments of top and bottom panels."

Thank you, this has been done as suggested.

10. *"Fig. 5: This is a key figure that was added in response to the 1st round of review to replace the now Table C. Yet, the main message — that DFO-LS requires much less evaluations than CMA-ES — is now obfuscated by the use of different scales and 2 y-axes. Better to show both on the same scale and let the visual speak for itself! The broken-axis suggestion (from the 1st review round) was not used, although it would make this much clearer in my opinion. Here is what I had in mind, e.g., for Fig. 5a (The red dashed line shows the imposed limit on evaluations for DFO-LS runs.): I understand MATLAB is not suited for broken-axis plots, so to be helpful I have provided below the python code that produces the broken-axis plot shown above. This code can easily be used as a template to reproduce each panel in Fig. 5. …"*

    We thank the reviewer for this comment, and for the time they spent to modify the Python code they have kindly provided. We have used this code to create each panel of Figure 5.