

We thank Referee 1 for their comments and help improving the manuscript. This document replies to those comments point by point, with R1's comments in green, our responses in black, additions to the manuscript in blue, and deletions in red.

Max Thomas et al. determine if and how a 1D parametrization of brine convection can be expanded to reproduce the laboratory PFA measurements of a previous study (Garnett et al. 2021). The paper consists of roughly four parts. In part one, the authors propose four different methods (A,B,C,D). In part two, they tune the free parameters of methods B, C, and D to minimize the absolute bias. In part three, they compare the model against the observed PFA profiles. Finally, in part four, they discuss if their results would apply to longer sea-ice simulations and other chemicals of interest.

The paper's text is clear and the method is clearly formulated (with one small exception) and applied thoroughly. The main conclusions are clearly stated, relevant to the field, and well supported by the results. The plots are mostly legible and not misleading, and the code has been made fully available. The topic of the submitted manuscript fits The Cryosphere. Based on my experience, the quality of the draft is well above average.

However, I found the structure of the paper's second half confusing, and the paper is somewhat ambiguous about its scope. Moreover, the figures could be improved upon, and there are a few minor other issues to address. Accordingly, I recommend accepting the submitted paper, but with minor revisions.

### Minor comments in roughly descending importance

- Scope. The introduction clearly states that the paper aims to determine if decoupling can explain the observed properties. However, the methods introduced and the results discussed go beyond that. I feel that one or two paragraphs are missing at the end of the introduction to describe the other main question of the paper, namely if the decoupling is linked to the surface area, brine salinity, or constant. Furthermore, here I feel that the expectations should be clearly stated. From the current draft, I am unsure which methods B, C, and D closest match the known theory.

Several changes to the text now better describe the scope.

We have added the following text to the introduction:

greater than that of the underlying water. It is plausible that  
15 PFAS were not perfectly dissolved in sea-ice brine during  
these experiments. The extreme cold and high salinity of  
brine has potential to drive solutes from solution, while the  
surface active properties of PFAS could cause them to stick  
to internal ice surfaces. Given that brine convection is the

And to the final introduction paragraph:

25 Here, we test that hypothesis by using observations of  
PFAS (Section 2.1) to evaluate a brine convection model  
(Section 2.2.1). We ran three simulations that include plau-  
sible mechanisms of decoupling and tune each mechanism  
against the observations (Section 2.2.2). Comparing the  
30 model performance to the observations allows us to test our  
hypothesis and identify the likely factors controlling PFAS  
dynamics in our experimental system (Section 3) ~~and to~~.  
Finally, in Section 4 we provide insights into how brine  
convection parameterisations can be adapted to model non-  
35 conservative chemicals in sea ice (~~Section 4~~).

We now describe more clearly how the good performance of Methods B and D are consistent with increased brine salinity driving PFAS out of solution:

similar dynamics. The gradient of the modelled concentra-  
40 tions using Method D regressed against the measurements  
is just inconsistent with 1 at 95 % confidence. However,  
given the improvement in model performance with Method  
D relative to Method A, and the similar performance be-  
tween Methods B (which ~~passed our tests~~ produced a gradient 45  
consistent with 1) and D, our results ~~do not rule out increased~~  
~~brine salinity~~ are consistent with increased  $S_{br}$  driving PFAS  
out of ~~solutions as solution as being~~ an important decou-  
pling mechanism in our study. Both methods are able to

(The poor performance of Method C suggests that surface area is not the controlling factor leading to non-conservative behaviour. We state this in the original manuscript and have retained it.)

We also revised the conclusions:

A sea-ice brine convection model can reproduce observations  
of poly- and perfluoroalkylated substances (PFAS) freezing  
into sea ice, providing a term is introduced that partially de-  
couples PFAS from the moving brine. PFAS with longer car-  
30 bon chains ~~behaved less conservatively were enriched~~ with  
respect to salinity ~~and required larger decoupling parameters~~  
~~—, sometimes to levels above the underlying water, in~~  
observations. Larger decoupling parameters were required to  
simulate this behaviour. Decoupling methods with primary  
35 dependence on the brine salinity performed well, and  
outperformed a decoupling method that depended on the  
internal sea-ice surface area. Our results demonstrate that  
brine convection models are powerful tools beyond pre-  
dicting sea-ice salinity. Complex, biogeochemical problems  
40 can benefit from the accurate, physically-based advection of  
brine predicted by current parameterisations.

- Structure. I missed the transitions between results, discussion, and conclusions on my first read. In my view, the tuning of the methods and the analysis of the resulting parameters are the first results. In the current draft, this is a single sentence at the end of 2.2, and is then revisited in Figure 4. Furthermore, I believe the results extend till line 173, and the discussion begins by discussing how general the results are. (I enjoyed the discussion along with the supplementary material.) From lines

129 to 173 I get lost between all the comparisons of B to C to D, and some things are repeated multiple times (e.g. lines 155-159). I recommend breaking down the results into more bite size chunks, answer a question and then move on to the next. One of the questions I would like to see answered is what it means that B and D are so similar. Is there no T dependence in reality? Or, is the data insufficient to distinguish?

On reflection, we retained Figure 1 and the text describing it in Section 2.2.2 (Methods of Decoupling). We feel that the parameter tuning is a technical detail, best placed there. We have revised the text to be clearer and have revised the figure caption to better introduce Figure 1 (please see new Figure 1 in response to the 3<sup>rd</sup> referee comment).

We have placed the beginning of the original discussion in the new results. The text has been reworked for clarity, and the similarity between methods B and D is now discussed (see below).

~~tests, where the~~ (Figure 3). Despite this, Methods B and D are similar in terms of profile shape and quantitative performance. This similarity arises because Method B also includes a (implicit) dependence on  $S_{br}$ . The product  $\gamma c_{si}/\phi$  (Equation 2) increases with  $S_{br}$  (for both methods) because  $\phi$  linearly scales with  $S_{br}$ . After tuning, both methods simulate similar dynamics. The gradient of the modelled concentrations using Method D regressed against the measurements is just inconsistent with 1 at 95 % confidence. However, given the improvement in model performance with Method D relative to Method A, and the similar performance between Methods B (which ~~passed our tests~~ produced a gradient consistent with 1) and D, our results ~~do not rule out increased brine salinity~~ are consistent with increased  $S_{br}$  driving PFAS out of ~~solutions as solution as~~ being an important decoupling mechanism in our study. Both methods are able to bring the model into reasonable agreement with the available observations and further measurements would be needed to robustly distinguish Methods B and D.

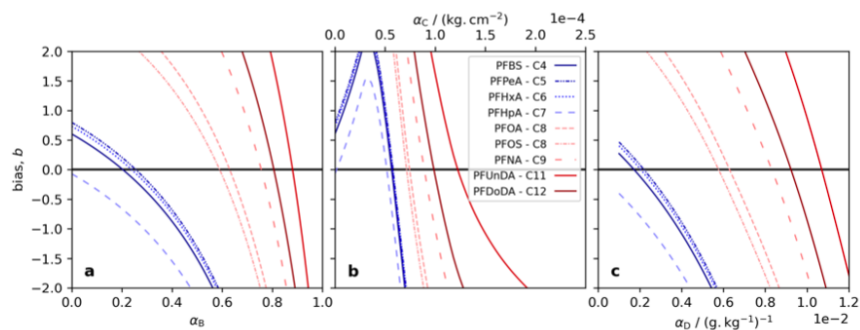
- What is the absolute bias  $|b|$ ? I assume it must be the absolute difference over the vertical sum of the modeled and measured concentrations. But in line 112, it says the difference between the measurements and the co-located model layers, which implies that  $|b|$  should only be zero when the model and obs match at all layers. Moreover, how is the absolute bias scaled? What does  $|b| = 1$  mean? Why use the absolute value? Showing  $b$  instead could clearly show that the higher alpha is, the higher the total concentration is. It might also make the lines in Subfigure 1b less confusing.

We have explained this better in the text, and describe (and plot) the bias rather than the absolute bias, which is more intuitive (see next comment). Please note that residuals are signed so positive bias in one region can compensate for negative bias in another (i.e. the model does not have to match exactly at all layers to achieve  $b=0$ ).

We choose  $\alpha$ , for methods B to D and for each PFAS, by running the model using 100 values of  $\alpha$  and selecting that which minimised the absolute bias between the measurements and picking the best performing  $\alpha$ . The performance was quantified using the bias,  $b$ , taken to be the sum of the residuals for each measurement relative to the co-located model layer depth at the end of the model run (Figure 1). simulation. The best performing  $\alpha$  gives  $b$  closest to 0. For each method, short-chained PFAS tend to produce lower  $\alpha$  and the two longest chained PFAS always give the highest  $\alpha$  (Table 1, Figure 1).

- Figure 1 has many lines that are difficult to distinguish. The readability could be improved by increasing the plots' width to use the paper's full width. The yellow line is also difficult to see; a darker tone would be helpful. There are no subfigure labels (a,b,c), and shifting the legend outside the area of the subfigure would also help. The current version, in which the legend blocks the lines' view and overlaps with the figure borders, is messy.

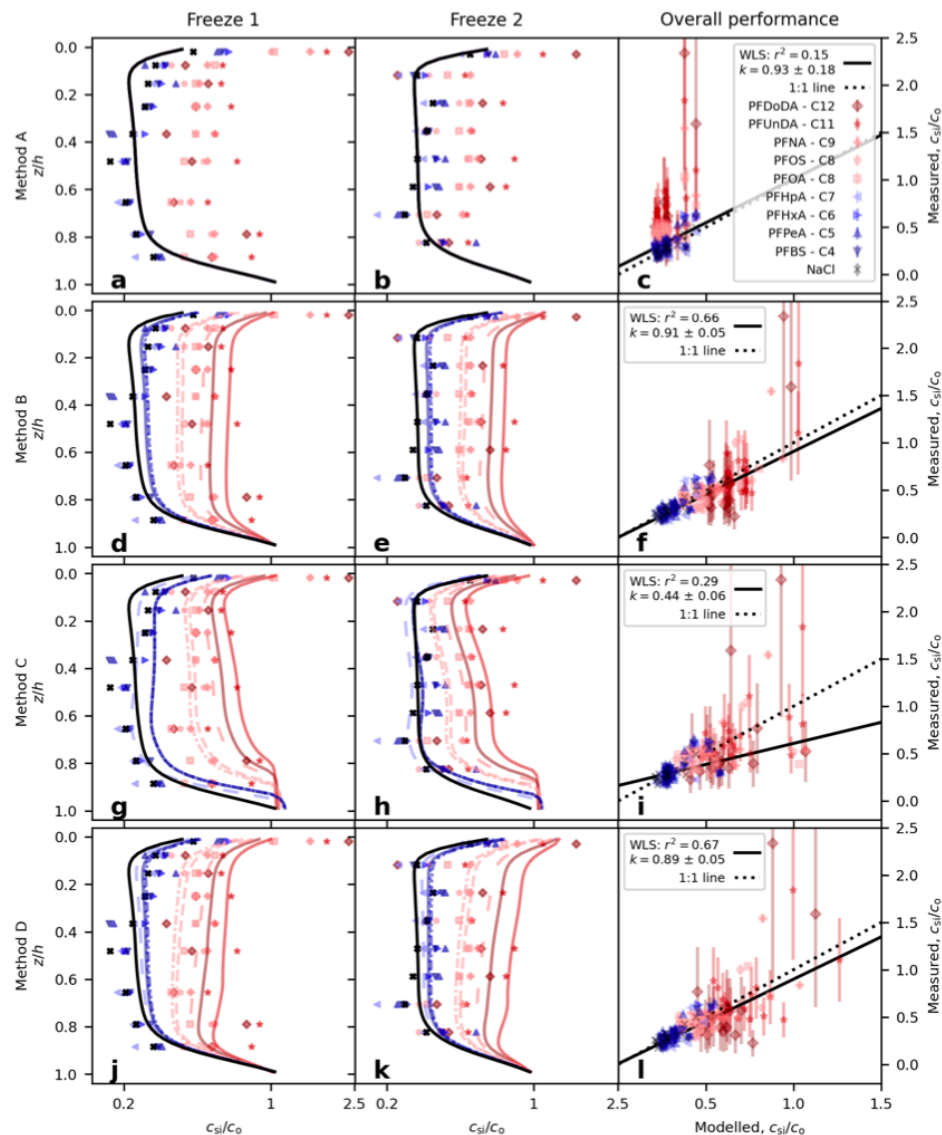
We have adjusted the colour scheme here (and for the rest of the manuscript). Rather than blue to yellow, the lines now go blue to red, which is good for colour blind readers and is clearer in general. Figure 1 now has a/b/c labels, the subplot spacing has been adjusted to maximise the space, and the figure spans more page width. The legend is now in panel b to avoid overlap with the lines. The lines are bolder. Another change is that we plot the bias, rather than the absolute bias, which is easier to interpret. The 0 bias line is added to highlight where the tuning parameters perform best. Some changes were made also in response to Referee 2. Please see the new figure below.



**Figure 1.** Absolute-Model tuning for each PFAS. The bias,  $b$ , was calculated as the sum of the difference between measurements and co-located model with a range of layers in three tuning parameters,  $\alpha$ , runs. Each panel shows results for Methods B a different decoupling method: a) simple decoupling, tuned using  $\alpha_B$ ; b) surface area adsorption, tuned using  $\alpha_C$ ; and c) salinity mediated decoupling, tuned using  $\alpha_D$ . The best performing  $\alpha$  give  $b$  closest to 0 (black line).

- Figure 2 has too many lines and markers in too little space. This figure could be separated into two figures for profiles and scatterplots, but at least make full use of the paper width to make columns 1 and 2 twice as wide. This is now a minor detail, but I was initially confused by the axis choice for the right column. Since they share the same observation data, it makes more sense that the observation data be the x-axis—shared data on the shared axis. For example, one could easily compare where the 2.5 measured C12 is in each plot.

We prefer to keep the scatter plots in this figure, which captures all the main results for the manuscript. To improve the readability, we have: adjusted the subplot spacing; adjusted the subplot grid so that the profiles are as wide as the scatter plots; adjusted the colour scheme (see above); reduced the marker size; and adjusted the limits of the axes. We will also use the full available paper width. With regards to the choice of x/y axis for the plots, we have retained our original choice because the dominant uncertainty should be on the y variable in a weighted least squares and it is more intuitive to show depth profiles with depth on the y axis. Please see the revised figure below.



**Figure 2.** Comparison of modelled and measured concentration for: Method A, perfectly dissolved chemicals (a, b, c); Method B, simple partitioning-decoupling (d, e, f); Method C, surface area adsorption (g, h, i); and Method D, salting-out-salinity mediated decoupling (j, k, l). Depth profiles are shown for freeze 1 (a, d, g, j) and freeze 2 (b, e, h, k). Modelled against measured concentrations are shown in panels c, f, i, and l, alongside the best fit weighted least squares (WLS) regression (black line, gradient  $k$  with one standard error and coefficient of determination  $r^2$  shown in legend) and the theoretical 1 to 1 line for perfect model behavior (dotted black). Concentration for the profiles is given on a log scale to highlight separation between the profiles.

- Lines 170 and 172 reference some tests that can be passed or failed. I have searched the submitted manuscript and find no clue what tests these are.

We have removed the mention of 'tests' and have instead said precisely what we mean:

similar dynamics. The gradient of the modelled concentrations using Method D regressed against the measurements is just inconsistent with 1 at 95 % confidence. However, 40

- "was not a useful method" line 155. "useful" is not a well-defined adjective in this context. I recommend stating that C is worse than B and D and better than A.

We have amended to say that C is worse than B and D. We have not said C performs better than A, as it is not obvious that it does in all respects, with the observations being more poorly represented near the lower interface for short chained PFAS than method A.

I am trying to understand why the authors chose the name method A instead of reference or control. It is not a flaw and does not need to change, but I did find it strange that the first "method of decoupling" is "none". Accordingly, there is alpha\_B, alpha\_C, and alpha\_D, but no alpha\_A, and so on.

We chose to label the experiments with no decoupling as Method A because this is the most obvious way to model these chemicals. Please note that we did model each chemical during Method A, and lines for each PFAS are plotted in the profiles, but they are exactly overlain by the NaCl. We revised the wording when introducing Method A.

~~Method A uses the same parameterisations as salt — so PFAS should behave identically to NaCl — models PFAS using the same numerical scheme as salinity, is the obvious choice for modelling solute dynamics, and serves as a quality control for other model scenarios (Methods B to D) involving decoupling mechanisms because PFAS should behave identically to NaCl.~~ 30 35