In this paper the authors use data from a measurement campaign and try to simulate a special case with three different microphysical schemes. They find that one of these schemes is better than the other two in reproducing the observations. This scheme is then further used to investigate a number of relations and dependencies between various quantities. The clouds are divided in three regimes and three forms of crystal habit are distinguished. The results are thus a bit complex and not easy to keep in mind.

To my view, this paper needs a major revision for the following reasons: The presentation of the goals, methods and results is weak. I found many places where a rewording is necessary. This is the major weakness, but it leads to unclarity in other respects. For instance, although it is indicated in the abstract that the main goal of the paper is to judge the ability of the three codes to reproduce the observations, that is that this is a paper about models, this impression is not corroborated later, where the impression is, the main goal is the investigation of the role of the different cloud regimes and ice habits in regulating the properties of the TTL. The reader would get more out of the paper if the findings could be presented in a more concise fashion, e.g. as a bullet list at the end. If you find a possibility to display your relations and sensitivities in a way that makes them physically plausible, it will even be better.

Another important issue is the statement (several occasions) that the observed large crystals do not fall. I have no idea how this could be observed with an airborne measurement. Instead, it seems that clouds observed at different locations and times have been compared to simulated clouds at one location at different times. It might be that the crystals do hardly fall in the model, because of very low density and mass, but I find it questionable to conclude from this model result to the reality which was observed in a different setting.

In the following I list some observations that I have made at reading, in the order of their occurrence.
1) The title is not good and I think, it is misleading. The abstract indicates that the paper is not about physics but rather about the performance of different ice cloud model versions in reproducing certain observations. The title should reflect this. Further it is unclear which sensitivity is meant: "Sensitivity of ... cirrus", what does this mean?

2) I also suggest to reorganise the abstract. The chain of arguments and statements does not form a logically coherent sequence, to my view. It starts with a general statement, then immediately jumps to the goal of the investigation, which however has to do with model details, not with something one could derive from the initial statement. Then it jumps again to reporting sentences on a field campaign. Please try to find a more coherent way to present 1) the general background, 2) the special problem, 3) how you treated it, and 4) what results you found. Also this text should reflect that you are investigating a certain detail in models rather than a physical phenomenon.

3) Language should be checked carefully. Line 45: "into" should be "in". Sentence beginning in line 55: "The impact ... is an ongoing debate". This is ugly style. I would rather say, the impact of ... is topic of a debate, but it is not a debate itself. Line 63 ff: "making them ... one of the largest uncertainties". No! the cirrus clouds are not uncertainties, but they may be the source of uncertainties. I stop now the collection of these kind of sloppy phrases and ask the authors to find and correct them themselves.

Lines 88-89: "We also investigate the impact of different ice habits on fall speed, relative humidity over ice and diabatic heating". Please reformulate. There are several issues in this sentence. First, I would not say that ice habits have an impact on fall speed (because there would be no fall speed if there were no crystals) but rather that their fall speed depends on their habit. Second, it is not clear to me how ice habit can impact on relative humidity, perhaps the sentence becomes correct, if cause and effect are reversed, i.e., the RHi impacts on the shape crystals attain.

Line 105: correct "number size distribution"

Figure 2: please add a scale, so that one can estimate the size of the crystals.

Line 129: is the model indeed compressible or does it rather use assumptions about compressible air?

Line 137: 900 km? or 900 m?
Line 160: please include the word "simulated" before TTL. Cirrus clouds are not sensitive to any parameterisations.

Lines 174-175 "but the growth of ice particle by vapor deposition remains constrained by mass–dimensional (m–D) and fall speed–dimensional (vt–D) relationships". I don’t understand this. I think that ice growth is proportional to supersaturation, i.e. depends on RHi, and that the prefactor depends on a number of other things like crystal shape, deposition coefficient etc. In my view the growth of the crystals is constrained by available supersaturation. Please explain, how the mentioned relationships interfere with this simple physics.

Figures 7-9: the observations are represented in these figures by a red line which extends over the whole indicated pressure range. As the Geophysica does not measure an IWC profile but just a value at a certain altitude, I suggest to replace the line with a red dot or a short red bar that can cover the two altitudes of the two HIWC measurements. The current version misleads to the interpretation that P3 performs better than Ishmael, but that seems not to match your interpretation.

Line 230 "are likely heterogeneously, or, less likely, homogeneously" what?

Text discussing Figs 10 and 11: Measurements show a lot of events at T<288K and around the green line, but simulations don’t. Instead, the Thompson scheme shows crystals at this low temperature at much higher IWC and with large ice number density; the P3 shows crystals with IWC at $10^4$ ppmv at T<188; the Ishmael show no events at this low T. Please provide more discussion about these findings. What is plausible, what not. And please, use "low temperature", not "cold temperature".

Text discussing Fig 14: To my view, this discussion can be improved. Currently there is just a description of the figures. I am, for instance, much interested to know why the so far best scheme shows so high supersaturation values. Do you think, they are realistic? What would happen, if there were indeed such high values of RHi, both in nature and in the model. Would you expect new nucleation (any mechanism)? Does the extreme supersaturation in the model explain the occurrence of the >800μm crystals? And why should high RHi favour an enhanced freeze drying process? To me, freeze drying means, that ice crystals consume all the available supersaturation at the coldest layer, such that only ice saturated vapour (RHi=100%) enters the stratosphere. If we however have huge RHi in a vertical uplift, freeze drying would have to act very quickly to get supersaturation consumed completely. This sounds implausible. So to my view, freeze-drying is not favoured at all, it simply gets unimportant at such conditions.

Line 266: how can a distribution be underestimated? Please reformulate.
Fig 15: As an effective radius is somehow the ratio of the third to the second moment of the size distribution, I think, it should be a single value. But here you show distributions. Do you consider the effective radii in every grid box, respectively, and then construct the histogram from these values? Please clarify. Why does the Ishmael scheme not produce larger crystals if the RH(\textit{i}) is so large in this model? Is this because sedimentation removes such crystals quickly? And finally, why do you show effective radius instead of a more simple measure like volume equivalent radius or maximum dimension?

Line 273: "lower RH(\textit{i}) saturation" is sloppy and unclear. Saturation means RH(\textit{i})=100\%. Or do you mean the corresponding vapour pressure?

Fig 16: It seems here that the model produces much larger crystals than observed, in line with the huge supersaturation. But in figure 15 we saw that the model has smaller effective radii than the observations. How can you explain that. It sounds like a contradiction.

Line 306 ff "During F8, it was reported that very large aggregates of ice particles were observed for an unusually extended long time without falling". How is this possible? Has the Geophysica made a stop there and somebody of its crew watched certain large ice crystals that were not falling? Or do you rather mean, you found large ice crystals at one place and a while later you found again large ice crystals, but at another place, and later. This would not imply the implausible interpretation that these crystals were not falling. Please explain more exactly what you mean. And then rethink your interpretation. The Geophysica observes different clouds or cloud parts at different times (the flight path does not show circling somewhere), but the simulations represent one cloud at one place at different times. Obviously these are different settings. How can you then relate the two. This sounds illogical.

Line 326: Perhaps the sentence becomes better, if you include "and thus indirectly affects ...".

Line 381: Please rewrite the sentence in a way that the reader does not conclude that ice in the TTL (\textit{i.e.} in the real atmosphere) is in any way dependent on details of any model simulations.

Line 381ff: please explain how the distributions have been obtained: are they taken over the whole model domain (each grid cell, but divided acc. to the cloud categories) and at each time step, or was the sampling different?

Section 4.5: I wonder what heating rate distributions for different habits and different cloud regimes mean in total, \textit{i.e.} which regime and which habit has the biggest effect. To
answer this question you would need to integrate the histograms and weight them with the respective frequencies of occurrence. Is that possible? It should tell us how many Gigajoules (or something similar) are stored in the three hours in the TTL, due to the presence of different crystal shapes in different cloud regimes. The mentioned doubling in heating rates could imply a considerable larger or smaller factor in the total energy deposited in three hours. I think, this is probably more relevant than the heating rate distributions.