Comments by Reviewer #1

We want to thank Reviewer #1 for their thoughtful and constructive comments that have improved the manuscript. Our responses below. We wrote our responses in the form of proposed changes to the text that we would make in a potential revised manuscript.

General: The manuscript presents a new, exiting interpretation of older and younger ice core elemental and isotope ratio results. I enjoyed reading it. The results of combined proxy, called 86Krex40, are based on relative difference of δ86Kr (86Kr/ 82Kr ratio values) and δ40Ar (40Ar/ 36Ar ratio values) to δ40Ar. It thus corresponds to a delta value of delta values expressed in permeg/permil when the primary delta values are expressed in permil. This relative double difference results in very small values and are therefore they expressed in permeg, which is a permil of permil. The measurements show that the values corrected for thermal diffusion are in the small negative range of 0 to -160 permeg/permil for 86Krex40. The authors state that 86Krex40 is a direct proxy of large-scale atmospheric circulation (synoptic-scale pressure variability). Yet, they are careful with their interpretation as there are still insufficient knowledge of the underlying firn air transport and gas trapping which may influence 86Kr ex40.

There are a couple of major points that should be addressed before the manuscript can be published.

Major points:

(1) Figure 3 is one of the major Figures and these values depend on two corrections applied (gas loss and thermal diffusion) which are detailed in Figure A3. Looking at the supplementary Figure A3 that displays the uncorrected and corrected values for gas loss and thermal diffusion independently, I saw that there must be an interdependence of these two corrections as they are not adding up. For instance for DF the uncorrected mean value is around 33 permeg/permil, the gas loss corrected about 55 permeg/permil which leads to a gas loss correction of around 22 permeg/permil. The correspondent thermal correction amounts to (33-(-10) = -43 permeg/permil both for individual or mean ΔT. This in combination would lead to correction of -21 permeg/permil (-43+22). I therefore would except overall corrected value of around 12 permeg/permil (33-21). The values plotted are, however close to +40 permeg/permil? Could you explain how and why they are interlinked, or is there a mistake for the DF values? The other site value corrections are more or less additive, maybe with the additional exception of GISP2.

Actually, the same issue concerns Figure A4.

We checked our calculations, and there is no mistake for the DF values. There is indeed an interdependence of these two corrections, and they are not additive. The observations of the reviewer are thus absolutely correct. The reason is that both involve the d40Ar isotopic ratio. The gas loss correction makes the d40Ar values smaller, which by itself makes $^{86}Kr_{xs}$ more positive at all sites. However, this change in the $\delta^{40}Ar$ also changes the estimated firn temperature gradient ΔT, because it is based on the 15N excess ($\delta^{15}N$-$\delta^{40}Ar$/4). Performing the thermal correction either with or without the gas loss correction therefore will give different results. There is therefore not a single value for the thermal correction, and the size of this correction is dependent on the size of the gas loss correction.

At none of the sites we expect the corrections to be exactly additive, however, depending on the details of the site they may appear approximately additive.

(2) There is hardly any information/discussion about the many more elemental and isotope ratio measurements that have been measured (section 2.2) to strengthen or weaken their arguments, i.e. 84Kr/ 86Kr, 84Kr being the most abundant and therefore the precision should be better.

Yes, ^{84}Kr is indeed more abundant, resulting in the largest signal on the IRMS cup. However, it has a smaller mass difference with the other isotopes. The $\delta^{86/82}$Kr has the largest mass difference (4 mass units), and we find that of all the isotope pairs it typically has the best precision per unit mass difference. For this reason, and for the sake of consistency, we decided to use this isotope pair throughout for the krypton isotopes.

In her PhD thesis, Anais Orsi introduced the concept of δ*Kr, which is the weighted average of all Krypton isotope pairs ($\delta^{86/82}$Kr/4, $\delta^{84/83}$Kr and $\delta^{86/84}$Kr/2), weighted by the standard deviation of the repeat measurements for each pair. For several of the sites we compared δ*Kr and δ^{86}Kr, and did not find much difference between these two. In future work we plan to investigate the difference more systematically.

(3) The expression of 86Kr ex40 being a direct proxy for synoptic-scale pressure variability comes at several places and is actually quite misleading as they correctly state that the gas measurements represent a time-averaged value. The average times are large (years to decades) compared to synoptic circulation events (days).

This is a good point. While we tried to make this point clearly in the text, it could easily be misunderstood. We considered using the term "pressure variance" instead of "pressure variability", but that has a mathematical meaning that is distinct from the way we define Φ. To address this point, we have replaced these statements with "time-averaged pressure variability" instead, to clarify that we cannot resolve individual storm systems. We also added a statement to the abstract to reflect this:

"The $^{86}Kr_{xs}$ reflects the time-averaged synoptic pressure variability over several years (site "storminess"), and does not record individual synoptic events."

(4) Calibration of the 86 Kr ex40 has been done with reanalysis data of the time range 1979 to 2017. This data show a large spatial variability in the Antarctic. However, whether the stability of the spatial calibration will hold for temporal interpretations is difficult to judge but this is certainly one major weakness. Yet, I see that it will be difficult to find arguments to support it.

This is indeed an important point, and we agree that our approach provides only the first-order proof. This issue has long plagued the interpretation of other ice core proxies, most notably the δ^{18}O of ice. Future efforts in climate modeling, combined with more observations of Kr-86 excess through time are needed to move beyond the spatial calibration. We address this point in the manuscript:

"The calibration of the $^{86}Kr_{xs}$ proxy is based on spatial regression. In applying the proxy relationship to temporal records, we make the implicit assumption that proxy behavior in the temporal and spatial dimensions is at least qualitatively similar. This assumption may prove incorrect. In particular, changes in insolation are known to impact firn microstructure and bubble close-off characteristics, which in turn impacts gas records of $\delta O_2/N_2$ and total air content (Bender, 2002; Raynaud et al., 2007). Since $^{86}Kr_{xs}$ is linked to the dispersivity of deep firn, it seems probable that insolation has a direct impact on $^{86}Kr_{xs}$ also via the firn microstructure. We will revisit this issue in our interpretation of the WDC $^{86}Kr_{xs}$ record (Section

5). Overall, we anticipate $^{86}Kr_{xs}$ to be a qualitative proxy for synoptic variability, yet want to caution against quantitative interpretation based on the spatial regression slope."

(5) In section 1.2, the authors discuss several processes that alter the isotope ratios, such as gravitational settling and thermal diffusion, advection, convection and dispersive mixing. The latter three they state do not distinguish between isotopologues. This is correct but they do lead to a disruption of the maybe already established isotope equilibrium through molecular (gravitational and/or thermal) diffusion, which requires time to be re-established. This is the starting point of their definition of the 86Krex40 proxy. However, there are several processes that can and will affect this proxy as nicely discussed in sections 1.2 and 3.3. This is of course a weakness of this proxy despite the author's transparent writing. For instance, they state that the major influence on 86Kr ex40 comes from pressure variations at the surface, but what about the pressure variations from the gas close-off process? Pressure variations may be weak but gas velocities of expelled molecules in the tiny channels at close-off depths might be very high and could lead to significant alterations in the gas compositions.

The reviewer is correct that pore closure from the firn densification process will drive an upward (macroscopic) air flux that will contribute to dispersive mixing throughout the firn column. To our knowledge, two previous studies have addressed this point. Both conclude that this effect is negligible compared to the barometric pumping driven by weather systems.

The first study is Schwander et al. 1988:

"The decreasing porosity during firnification also leads to an air flow in the firn. When the firn is compacted, air is expelled from the open-pore volume. This leads to an upward movement of the air relative to the firn. The corresponding mean air velocity is of the order of the snow-accumulation rate, which is generally less than 10^{-4} mm/s. In the case of firn without melt layers (uniform upward flow relative to the firn at a given depth level), the flow-dependent part of the diffusivity is again negligible."

The second study is Buizert and Severinghaus 2016:

"Another source of macroscopic air movement in deep firn is the gradual closure of the pore space by the densification process, which leads to an upward air flow relative to the firn matrix (Rommelaere et al., 1997). The velocity of this (accumulation-rate-dependent) back flux is of the order of 10^{-9} to 10^{-8} m s^{-1}, and is clearly negligible in magnitude compared to the barometrically driven flow."

To address this issue, we added the following sentence to section 1.2:

"The upward air flow due to gradual pore closure is orders of magnitude smaller than the flows driven via barometric pumping, and neglected here."

Minor points:

L144: it would be worthwhile to explain why this definition is less sensitive to thermal diffusion (give corresponding reference). Yet,

We changed to: *"The $^{86}Kr_{xs40}$ definition is preferred, because per unit mass difference $\delta^{40}Ar$ is less sensitive to thermal fractionation than $\delta^{15}N$ is (Grachev and Severinghaus, 2003a; 2003b)"*

L144ff: What about close-off fractionation? We know that Ne, He, will expelled during close-off. Therefore, large molecules such as N2 will be less affected than O2, Ar, etc. Kr is obviously between N2 and O2. Especially, Ar will be subject of expelling. In this regard, the second definition with δ15N would be preferred.

Out of the three gases (Ar, Kr, N_2) Ar is the only one that is impacted by close-off size fractionation. This definitely impacts the $\delta Ar/N_2$ ratios which are negative (Fig. A1B). Fortunately, $\delta^{40}Ar$ is impacted only weakly by gas loss during close-off fractionation. We discuss this impact in appendix A1.

Note that both definitions are impacted by the $\delta^{40}Ar$ gas loss correction, as $\delta^{40}Ar$ is also needed for the thermal correction.

The observations make it clear that the $^{86}Kr_{xs40}$ definition has less scatter than the $^{86}Kr_{xs15}$ definition – compare Figs. A3 and A4.

L225ff: I would prefer the gas splitting. As it can be tested by many gas species measurements in contrast to the different ice core samples. There only replications can help.

Yes, so do we. It is more time consuming, though. In future work we plan to only use the gas splitting approach.

L239f How extended is this bubble-clathrate zone as the signals are extremely small. What was the criteria for the given number in the depth or time range.

We based our choice of the BCTZ depth/age range on observed positive anomalies in dO2/N2 in WDC ice, which occurred between 1000 and 1600 m depth.

L242ff "Some of the EDC samples analyzed had clear evidence of drill liquid contamination, which acts to artefactually lower 86 Kr ex40; the late Holocene data used here were not flagged for drill liquid contamination." Give a reference for this statement.

We added a reference to (Baggenstos et al., 2019). We did not make the determination of drill liquid contamination, but relied on the original study for this observation. We now also explain that d86Kr excess is lowered via isobaric interference on mass 82.

L247 22 per meg /permil: I do not understand this, error propagation leads to a higher combined uncertainty and not a lower!, this does not make sense or do I miss something here?

Good question. This depends on the value of the denominator, and for the values given in the paper we assume a $\delta^{40}Ar$ of around 1.2 permil that is typical for WAIS Divide.

Starting from the definition of Eq (2), the uncertainty in the numerator is effectively equal to the uncertainty of the $\delta^{86}Kr$ measurement. Because the value of the denominator is typically greater than 1, the uncertainty of the $^{86}Kr_{xs}$ appears smaller than that of $\delta^{86}Kr$ (in the WDC example, 26/1.2 = 22). Of course in a relative sense the $^{86}Kr_{xs}$ error is much greater than the $\delta^{86}Kr$ error.

We now clarify this in the text: *"Via standard error propagation, this results in a ~ 22 per meg ‰$^{-1}$ (2σ) analytical uncertainty for both $^{86}Kr_{xs40}$ and $^{86}Kr_{xs15}$ at a site like WDC where $\delta^{40}Ar \approx 1.2$ ‰."*

L248 this would be necessary. At least you can split a 1600 g samples in two sub-samples

Yes, this would be a good thing to try in future studies. For traditional ice cores it is challenging to get a 1600 g sample, however, as we typically do not have access to the full core but only the gas piece. It would further lead to very skinny long samples with a lot of exposed surface areas, which is not ideal as this may result in gas loss during pumping. For blue ice sites (Taylor glacier, Allan Hills) we tend to have much larger samples available, yet at such sites the orientation of the stratigraphy is poorly known and true depth replicates are rarely true age-replicates. We will consider this for future work.

L253 BP also denotes Before Present, consider changing it.

Good idea. We changed it throughout to BRP.

L257ff Are there additional indications that melting has occurred, for instance from water isotopes or changed greenhouse gas concentrations?

Possibly. We have not made such measurements. It would presumably result in elevated CH_4 and CO_2. We are unsure what the impact on water isotopes would be. In the absence of the greenhouse gas measurements we rely on the noble gases, which we believe are a sensitive indicator for melt.

L268 Can you specify what modern climate means (time range)

In this statement we rely on the ERA interim reanalysis dataset we used, which is from 1979-2017. We now specify this:

"pressure variability in the modern climate (here: 1979-2017 CE)."

L274 How can a daily variable be compared to a decadal variable (86Kr ex40)?

In all our analyses we compare Kr-86 excess to the multi-decadal average Φ (1979-2017 CE). We now specify this more clearly:

"In Fig. 3A we plot the site mean 86Krxs40 (with ±1σ error bars) as a function of Φ (averaged over full 1979-2017 period)"

L278f If such a calibration is made, it should be done on firn air samples as they are not smoothed by the process of gas enclosure. Have you tried to do this?

We do have very limited firn air data for $\delta^{86}Kr$, and we do not have access to stored firn air from a wide range of climatic conditions. So unfortunately, we cannot perform this analysis on firn air. Furthermore, we find that firn air $^{86}Kr_{xs}$ does not match the values measured in mature ice samples. We have no clear explanation for this mismatch. We discuss this in section 3.3:

"Measurements on firn air samples, where available, suggest a smaller 86Krxs anomaly in firn air than found in ice core samples from the same site. We attribute this in part to a seasonal bias that is introduced by the fact that firn air sampling always takes place during the summer months, whereas the synoptic variability that drives the Kr-86 excess anomalies is largest during the winter (Fig. 2C); consequently, firn air observations are biased towards weaker 86Krxs. Further, in the deep firn where 86Krxs anomalies are largest, firn air pumping may not yield a representative air sample, but rather be biased towards the well-connected porosity at the expense of poorly-connected cul-de-sac-like pore clusters. Since barometric pumping ventilates this well-connected porespace with low-86Krxs air from shallower depths, the firn air sampling may not capture a representative 86Krxs value of the full firn air

content. These explanations are all somewhat speculative, and a definitive understanding of the firn-ice differences is lacking at this stage."

L288f This is only the case when the argon correction for gas loss has been made correctly.

They would be identical for any value of ε_{40}. But the result indeed relies on the argon gas loss correction. We believe we have been forthright about this, and dedicated several figures and an appendix to this correction (Figs 3B, A3, A4).

L299ff Refer to Figure 3.

We added a reference to Fig. 3A as suggested.

L304ff This is disconnected, the link is not clear. Further explanation is needed here.

Thanks for this feedback, it is indeed probably not very clear to readers unfamiliar with the cited Koutavas paper. We have instead removed this sentence, as it is not important to the paper (which is long enough as it is!).

L310ff This is indeed a critical point.

Yes, it really is. We believe all the points in section 3.3 are critical, which is why we try to be very cautious in our writing and be clear that the results are tentative at this point.

L332f This is again a critical point as I can imagine that the firn structure acts as a column retarding the gas species differently. It would be worthwhile to do such experiments. Maybe you find a corresponding reference?

Agreed. We are considering writing a follow-up proposal to do such experiments. For now we can only speculate, unfortunately.

L346ff This is also important. Elemental ratio should be in line with isotope ratios. Yet it points indeed to a difference in diffusion coefficient ratio of real and lab conditions. Column effect (adsorption/desorption)

Agreed. We edited this paragraph in revision to include the adsorption effect:

"Firn models predict that the gravitational disequilibrium effect in elemental ratios (such as 𝛿Kr/Ar) should be proportional to that in isotopic ratios. However, the observations suggest that the former is usually smaller than would be expected from the latter. As before, adsorption of Kr onto firn grain surfaces may contribute to the observed discrepancy, and laboratory tests of this process are called for. Further, the impacts of gas loss are greater on elemental ratios than on the isotopic ratios which may contribute also. Including measurements of xenon isotopes and elemental ratios in future measurement campaigns may be able to provide additional constraints to better understand this discrepancy."

L361f if one argues that the diffusion of noble gases may be retarded in the firn column, one should consider this effect also for the thermal corrections. These, however is based on δ15 N and δ40 Ar measurements.

The estimated diffusivities of ^{15}N and Ar are very similar (Buizert et al., 2012), so therefore we ignore this to first order. The effect of the firn column on retarding these via adsorption are not well known. Ar

has a higher solubility, and therefore is likely to have a stronger adsorption effect. However, the adsorption for both these gases is estimated to be small compared to Kr and Xe though.

L368ff this is also a very critical correction as obliviously ΔT varies considerably from site to site without a clear understanding why this is the case.

Yes, we agree. The reviewer does not seem to be asking for a correction or response here.

L378 give a reference for (1)

We added the reference to Baggenstos et al., (2019)

L379 (3) yes, this indicates the large uncertainty of this correction. Yet, Figure 3B is quite convincing as a counter-argument. See also main comments above.

Agreed. We responded to the main comments above, no further response seems needed here.

L 410 How has the elevation changed over the course of the investigated period? And how relevant are these changes?

This is not well known, and we need to rely on models here. For WDC, Golledge et al. (2014) simulate an LGM elevation of around 300 m higher than at present. Via the regression slope, this would result in a 10 per meg $‰^{-1}$ change in $^{86}Kr_{xs}$. This is within the uncertainty of our measurements, and smaller than the temporal signals we interpret.

L416 ...(by limiting ...), not clear, needs further explanation

We clarified this statement to read: *"via its topographic influence on the position of storm tracks"*

L429f ...we anticipate 86Kr ex40 to be a qualitative proxy for synoptic variability... this is indeed a good point as the used calibration is standing on weak grounds.

The reviewer does not seem to be asking for a correction or response here.

L456ff Why have you only investigated Antarctic sites and not Greenland locations? This would proof that different locations on Earth would be similarly influenced. There is GISP2. What would be a good choice for additional stations in Greenland?

A map of synoptic variability in Greenland is given in figure 8 of Buizert and Severinghaus (2016). It is clear that Greenland has only a small range of barometric variability compared to Antarctica. The main coring sites in the interior all have a comparable level of pressure variability, making them uninteresting for the spatial calibration. Cores near the margin, such as the Renland core, have greater pressure variability, yet nothing near the levels seen in coastal Antarctica. Greenland coastal sites like Renland suffer from summer melting, however, which impacts the Kr and Xe inventory of the ice strongly. For this reason we decided to use only a single Greenland core in this study. In the future it would be interesting to obtain a long-term record of Kr-86 excess from Greenland summit.

L549f or is there indeed a higher variability present. How do you explain or underline that less care has been taken for these later campaigns?

These measurement campaigns were aimed at obtaining a $\delta Kr/N_2$ record for reconstructing mean ocean temperature, and Kr-86 excess was a by-product of those measurements. For these reasons less care was given to the isotopes than desired. Perhaps the experimenters were looking to complete the project on a short time schedule.

L575f Campaign 3 data shows quite a large scatter.

- We have added this caveat to the text: *"The trend in campaign 3 is less robust due to the greater scatter in the data"*

L600ff this might be tackled with measurements in Greenland compared to those in Antarctica

This is a great idea for future studies.