We thank the reviewer for the comments. As a general note we would like to stress that this work was primarily motivated by an observation of a spurious long-term trend in the derived volumetric water content (VWC) at some of the wetter sites in the COSMOS-UK network, for instance Glensaugh. It was identified that this trend was correlated with the Jungfraujoch counts used in the background neutron intensity correction, then of the form $I(t)/I_{\text{ref}}$ (so $G = 1$ in eqn. 8). We found that after applying other reported approaches (Hawdon et al., 2014), the long-term trend would still remain. So we took a data-centric approach and directly calibrated the normalised fluctuations of Jungfraujoch counts with the CRNS counts at the specific site of interest to find $G$ for that specific site. We also took caution to not fit through soil moisture variations by applying a lower quartile regression when deriving our site specific $G$s. Our method significantly reduced the observed long-term trend.

We of course appreciate the reviewer’s request for clarity, providing comparison of approaches and more scientific understanding of how $G$ varies with respect to different physical processes. We will address these points in the new version of the manuscript. We also apologise for the misstatements in terms of what is revised: the form of the correction given in eqn. 8 appears in the literature (although with almost no explanation where it originates from) and it is the method for obtaining the site specific $G$s that is revised. We will clarify these aspects in our manuscript.

Below are our responses to the individual comments.

1. "The authors claim to derive a "revised correction factor" using an amplitude scaling technique that has already been proposed in the literature before, while suggesting the use of median instead of reference count rates. In the same study, the authors find that the proposed median count rate often is insignificantly different from the reference count rate, that it turns the whole approach to be highly dependent on the individual measurement period, and that it is less generalizable.”

We agree that the general form of the correction in eqn. 8 is not novel and has been proposed in the literature. The value of the paper comes not in the general form of the correction, but in the way we estimate $G$ – using local cosmic ray neutron sensor data
combined with a lower quartile regression fitting to fit through wetter points associated with more constant soil moisture in the UK. We apologise for the misstatements in the manuscript, and we will revise it to clarify this point.

To address the point of poorer generalizability due to the individual measurement period dependence, the following are two reasons why such dependence arises:

- **G** values are derived by comparing neutron count fluctuations from the reference neutron monitor (\(I_n\)) and local neutron fluctuations from the Cosmic Ray Neutron Sensor (\(N_n\)) and therefore there is a necessity to have sufficient variations in \(I_n\) to establish a reliable relationship between \(I_n\) and \(N_n\). This is linked with the solar activity more than the time-period – it is already explained in the manuscript in L 141 and L 142, but we will clarify this further. Small variations in \(I_n\) indicate low solar activity in which case the intensity correction is of less significance. So while the method requires sufficient variations in the incoming background neutrons, it is also mostly not required when this is not the case.

- The CRNS data carry soil moisture signal and therefore the measurement period needs to be sufficiently long so that it is possible to fit through relatively constant soil moisture points (in the case of the UK, these are the points associated with wet periods). So it is rightly noted that for a short time series \(G\) will carry more uncertainty, but since the method is aimed at correcting long-term trends, the users requiring this method will most likely have a long time series available.

2. "**Good agreement has been found by comparing the overall average at tens of sites with findings from other studies, but the reader is still left with no clear idea about the added value provided by this study.**"

The original motivation for this study was an observation of a spurious long-term drying trend present at wetter COSMOS-UK sites, for instance Glensaugh, when \(G = 1\). Using \(G = 1.22\) based on cut-off rigidities, as given by Hawdon et al., 2014, also does not remove this trend completely. The added value of our study is that we calculate \(G\) for each individual COSMOS-UK site using directly neutron data from the local site of interest. While these neutrons are subject to soil moisture variability, we use a lower quartile regression fitting to put more weight on data points associated with more constant soil moisture. This approach corrects the observed unphysical drying trend present at some of the COSMOS-UK sites.

In the study conducted by Howat et al., 2018, a value of \(gamma = 1.19\) is given for a specific location in Greenland and the authors use a neutron monitor (THUL) also located in Greenland. \(Gamma\) given in that work is therefore for that specific case study and is not expected to be the same at different locations. While we note in the manuscript that our COSMOS-UK site average \(G\) value coincides closely with the value of \(gamma\) obtained by Howat et al., 2018, we do not use it in the background neutron intensity correction. We always use the site specific value which can vary significantly from site to site (between \(G = 1\) and \(G = 1.5\)).

We will improve the manuscript to stress more clearly the novelty of our approach and its implications on the derived soil moisture data. We will also add to the scientific understanding of how \(G\) varies with different physical parameters.

3. "**what is the expected sensitvity of CRNS to these corrections (cite e.g. Baroni et al. 2018 (JoH))**"

We will report on the sensitivity of CRNS to the background neutron intensity correction factor based on the literature.
4. “provide fundamental derivation of your approach and discuss differences to previous approaches, and”

If the reviewer is asking about how we derive G, then yes, we will clarify this further in the manuscript. We will also discuss the differences to other approaches.

5. “statistically sound evaluation (using all your individual site data, proper statistical measures, and uncertainty analysis)"

We will provide uncertainty estimates for individual site Gs.

Regarding the aspect of “using all your site data”, we did use all site data for the site specific G estimation. If what is meant is spatial analysis, then we will analyse our data in the context of understanding why G varies from site to site.

6. “The fundamental explanation of the involved processes and physics have been marginally addressed. The authors seem to treat the correction approach as yet-another function without discussing the meaning of the analytical form or the physical processes involved.”

We took a data-centric approach without complex modelling of the physical processes involved, but we will improve the manuscript by discussing the physical processes.

7. “Why should detected low-energy neutrons in any way behave proportional to incoming high-energy neutrons? What fraction of the detected signal is direct and indirect radiation and how does this change with site conditions? The authors you are citing have answers to that. This would help readers to actually understand your approach and this may have direct implications on the interpretation of the performance of your results at individual sites.”

These of course are very valid questions, which we had originally decided to defer to later work, but we now see that this work is somewhat incomplete without at least discussion of the relevant physical processes and we agree that these would help to inform the interpretation of the different G values across all 48 sites.

We will discuss the energy spectral response functions for the CRNS versus the RNMS with respect to the question of proportionality between the incoming low and high-energy neutrons, and the relevance of site conditions for the correction. As above, will discuss properly the literature on this topic.

8. “Can we expect this approach to be different for different regions on Earth? What has this to do with latitude or cut-off rigidity? The answers may have implications on the performance of your distributed sensors and should be discussed using the empirically obtained values for G.”

We will show the dependency that our empirical G values have on cut-off rigidity.

9. “What could be the physical reason for the necessity of an amplitude scaling in the form of G*(I/I_m-1)? This has implications on the choice of G and its meaning, and could answer whether situations could occur with G<1, too.”

We will discuss the form of the correction and try to elaborate why this form can be adopted.

10. “Also put the whole idea into context and note that it is about correcting for solar activity fluctuations and mention that the scale of those variations can be
from years (solar cycle) to days (FDs, GLEs). This has direct implications on the conclusions that you could and could not draw from your analysis, e.g. by choosing a "monthly average", with which one would have no chance to resolve any short-term variations.”

In our study, we are concerned with the long-term trend (the solar cycle) that was observed at some of the wetter sites in the COSMOS-UK network, for instance Glensaugh. So we choose monthly averages in figures 3 and 4 to show the impact of the correction on long-term trends.

11. “The authors determine gamma empirically, while there are high chances of fitting through soil moisture variations. A more "safe" approach would be to calibrate gamma on quite periods (low solar activity) and at sites with constant soil moisture (desert, concrete, or lakes). Some authors (which you have cited) have already traveled along this path and it could be helpful to report on advantages and disadvantages of the various approaches.”

To prevent from fitting through wide soil moisture variations, we use a lower quartile regression when comparing the normalised neutron count fluctuations from the cosmic-ray neutron sensor (\(N_n\)) and the reference neutron monitor (\(I_n\)) to obtain \(G\). This ensures that we can use all the data, but the fitting is biased towards wet periods which have more constant soil moisture in the UK. We will stress it more clearly in the manuscript.

We will emphasise more clearly that we can only get good \(G\) estimates, over a sufficient change in RNMS counts e.g. with significant change in solar activity.

Some of the co-authors have also travelled this path to a degree with COSMOS-Rover experiments. However, given that cut-off rigidity plays a role, we do not have convenient constant soil moisture sites close to the SMS locations. There are a few exceptions, but even translating results from concrete or 100% water (lake) presents issues. We will discuss these approaches along with other onsite reference measurement approaches.

12. “The authors spend two pages to derive their concept, and end up with an equation which is equal to the one from literature which has been presented already in the introduction. It is claimed that the new equation is now "revised", for two reasons:

- it uses median count rates instead of a reference count rate (although this option has already been suggested by Zreda et al. 2012 and others).
  Interestingly, in your own study you tell that the difference between the use of \(I_n\) or \(I_{\text{ref}}\) often is insignificant...
- the scaling factor \(G\) now is empirical (although this has been done already by other authors which are cited elsewhere in this very study), while they are hardly compared to values from existing approaches. In addition to that, even the relationships to the cut-off rigidity, as proposed by Hawdon 2014, for instance, are empirical. So instead of empirically building on their theory, you reduce the complexity of finding \(G\) even more by applying a purely empirical fit. This comes with the risk of low transferability and generalizability of the results.”

The revision comes in the way we calculate \(G\) using the site specific CRNS data. Some previous authors used neutron monitor data (different detector and not site specific to where the cosmic-ray neutron sensor is located) to relate \(\text{gamma}\) (or \(G\)) to other physical parameters (Hawdon et al., 2014 relates \(\text{gamma}\) to cut-off rigidity and Howat et al., 2018 relates \(\text{gamma}\) to elevation, but with no formula given in the latter case).
While both approaches are empirical, our approach directly finds $G$ for the site of interest using that same site data and without making assumptions of what physical parameters are involved – we directly use the CRNS data from the site of interest and account for soil moisture variability. Using the CRNS data has also the advantage of calibrating $G$ for detector differences that exist between the CRNS and the reference neutron monitor, even if it is only a first order approximation.

We discuss the generalizability aspect in point 1 of this response and of course parameterisation of $G$ (if all relevant factors are included) would increase the generalizability of the results and improve scientific understanding of the processes involved. We will address this point by providing a discussion on the physical influences on $G$. We will also attempt to build on the Hawdon et al., 2014 approach for $G$ to explain our results at the COSMOS-UK sites.

We would also like to note that the work presented here has a very pragmatic focus: to bias correct long-term UK soil moisture time series data, for use by the hydrological and land surface climate model communities. We welcome others to test the applicability to other regions, to bias correct their historical data. The results will be specific to the make and model of the CRNS, and to a degree the installation conditions, as well as the surrounding environment.

13. “Under these circumstances, I strongly suggest to properly discuss previous literature and not to sell the current development as a "revised correction approach". This study for me looks rather like a case study, where some of the existing suggestions have been tested on the UK network data. It is surely important to do that, but it would require a different storyline.”

We apologise for the unintentional mis-selling – and we agree that we present an empirical application of the earlier published correction. We do consider that we present important details of the exact application, to allow an empirical fit using the on-site CRNS data (without underlying assumption of the causes) and to weight our fits to wet data to find $G$. We will strengthen the paper, as you suggest to provide more analysis of the results and more insight, as a UK case study.

14. “L143: If the results are dependent on the variation of the incoming radiation, then it should be easy to evaluate this by testing the approach during high and low solar activity. Something that could be done at the COSMOS-UK Network.”

We perhaps should have made it clearer: since $G$ is a slope of a linear regression (lower quartile regression) between $x$ ($I_n$) and $y$ ($N_n$) there needs to be sufficient variations in $x$ to find a reliable estimate of the slope. Figure 5c shows how the T values of $G$ estimates for the COSMOS-UK sites increase when the variation in Jungfraujoch counts increase.

15. “L147: "Preliminary studies show no simple relationship with ... latitude, altitude, or rigidity". In the last sentence before the conclusions you are mentioning the most important aspect, in my view, which should be central to this study. Consider elaborating on building a theoretical basis for your approach, which involves confrontation with the mentioned quantities. This would help to develop this approach further towards a generalized and transferable method.”

Agreed, as in earlier reviewer responses, we will improve this work by properly discussing the physical processes that may contribute to the fitted values.

16. “Common and vague language is used quite often, rather that clear and
factual language. E.g., L17 "perhaps less obvious", L85 "gives a gradient G, not of 1, but 1.5", L122: "change is very much greater".

We will improve the quality of the language used.

17. "L33: the Desilets equation is semi-empirical, the parameters have been derived from observations only. In general, this paragraph can be much shorter. The choice of the conversion function and vegetation or organic carbon corrections are of lesser importance here."

According to Desilets et al., 2010, the calibration curve was obtained by “fitting ground-level neutron fluxes simulated in MCNPX [Pelowitz, 2005] to the shape-defining function”, so we are not sure why the reviewer suggests that “the parameters have been derived from observations only”. And yes, we can shorten this paragraph.

18. "Fig 3: The comparison of time serieses with different G is not insightful. Please provide reference data, e.g. from the neutron monitor (to demonstrate a changing correlation to their signals) or soil moisture data (to demonstrate better match)."

We will revise the plot and add neutron monitor data to it.

19. "Fig 4: Now you show TDT data, but only with G=1, so there is no chance to compare the approaches and to evaluate the performance of your idea. Moreover, consider showing a weighted average using different depths, instead of only 10 cm data."

Yes, we will add a plot with the revised G to figure 4. Unfortunately, we only have TDT data for a depth of 10 cm, so we will not be able to show a weighted average.