Author Response to Reviews of

Improving Calibration and Validation of Cosmic-Ray Neutron Sensors in the Light of Spatial Sensitivity – Theory and Evidence

M. Schrön, M. Köhli, L. Scheiffele, J. Iwema, H. R. Bogena, L. Lv, E. Martini, G. Baroni, R. Rosolem, J. Weimar, J. Mai, M. Cuntz, C. Rebmann, S. E. Oswald, P. Dietrich, U. Schmidt, S. Zacharias
Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-148

RC: ***Reviewer Comment***, AR: *Author Response*, ☐ Manuscript text

Dear Prof. Dr. Bob Su,

we are greatful to the reviewers for the helpful comments and suggestions. We are addressing the reviews and the corresponding changes individually in sections 1–3 of this document. Apart from that, we would like to ask for the permission to include a few additional changes in section 4 of this document. The revised manuscript has been attached to this letter, together with a document showing the complete list of changes.

1. Reviewer #1

1.1. Major concern #1

RC: ***The author claimed that the revised weighting approach has the potential to reveal otherwise invisible hydrological features, like water ponding in remote or local areas, water in the biomass or litter layer, interception water storage, groundwater rise etc. On the other hand, such claim is based on a simplified interpretation by using a lump-sum expression of "excess water storage" in Figure 8. This renders the statement on a weak ground. Unless there are other experimental results or numerical simulation results can provide further proofs for this statement, I am not convinced with this statement by only looking at Figure 8.***

AR: *Thank you for pointing this out. The CRNS indeed is an integral signal which is only able to sense "excess water storages" when combined with independent soil moisture measurements. From the dynamics of that signal (e.g., following rain events) and additional supporting observations, it is then possible to infer hydrological processes, such as interception or ponding. This was suggested by Desilets et al (2010) and it was used in Baroni and Oswald (2015). In the latter study, seasonal overestimation was attributed to growth biomass. This signal was then removed. The remaining subdaily residuals were analyzed and attributed to water interception and by identifying also a daily cycle.*

In our study, ponding in the Großes Bruch *was actually observed independently during these days (see Fig. 1 in this very document). In the case of* Wüstebach*, Bogena et al. (2013) quantified water in the litter layer and observed ponding all over the catchment. To further support these interpretations, Figs. 7 and 8 (now 8 and 9) included information about groundwater rise in the same periods.*

While the mentioned additional information gave us evidence about additional hydrogen pools, we agree that the CRNS cannot distinguish the different pools. We softened that statement throughout the manuscript, using you suggestion of "excess water storages". See e.g., caption of Fig. 8 (now 9):

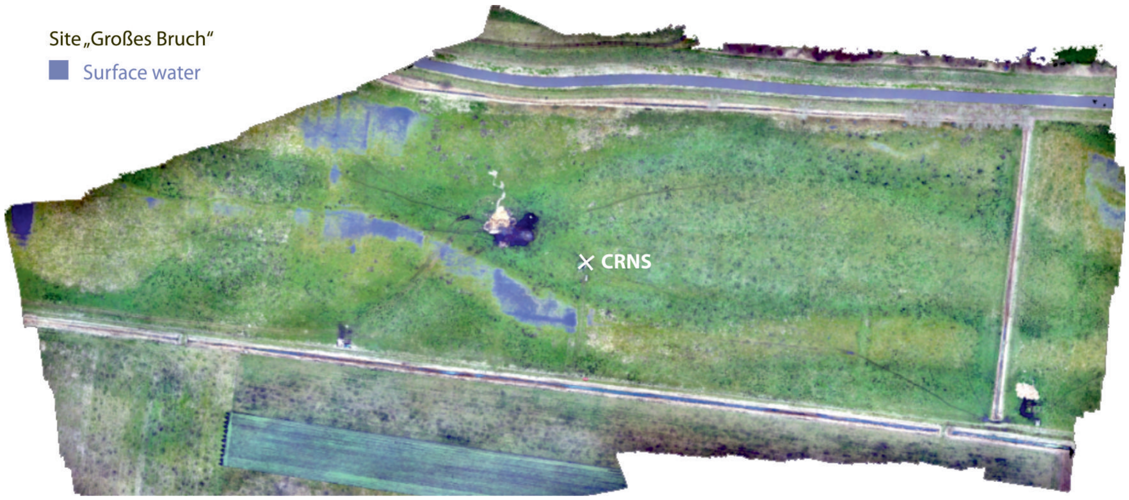

Figure 1: (not in the paper) UAV-derived orthophoto showing ponded water at the site *Großes Bruch*.

> Fitting the CRNS data to the SoilNet *conventional* average almost completely hides effects ~~like~~ from excess water storages (which could include water in the litter ~~and~~ layer, interception water, groundwater rise, or ponding close to the stream). The *revised* approach emphasizes those additional hydrological processes, while still robustly increasing the sensor performance.

And the conclusions:

> When CRNS is combined with independent soil moisture measurements, the revised weighting approach has the potential to reveal ~~otherwise invisible hydrological features , like~~ hydrological features that were otherwise lost in the signal by overcalibration. The approach improved the accuracy by which the CRNS probe was able to sense total changes of water storages other than soil moisture, e.g., from water in the biomass or litter layer, interception water~~storage~~, groundwater rise, as well as ponding in remote or local areas.

RC: ***Furthermore, the area difference between the "CRNS fit" and "revised" curve has two parts: one part is above the "revised" curve, and was defined as "excess water storage"; the other part is below the "revised" curve. However, for this "below part" there is no any explanation.***

AR: *This is a good question. As can be seen in the plot, the light blue curve goes below the blue curve many times throughout the period, and the maximum deviation always has a similar value ($\Delta\theta_v \approx 5\,\%$). We think that in these (dry) periods the CRNS signal mainly reflects the actual soil moisture, while in the other periods excess water storages are present. Thus, if the CRNS would be calibrated only on the dry periods, the lines would match very well, and most deviations would be positive (towards the wetter state), reflecting excess water storages. However, for the sake of comparability, we show only the calibration on the full period, as was done also for other sites. It is remarkable that even in this case better performance plus identification of excess water storages has been achieved using the revised approach. When the dataset would be used in* *future studies to identify those hydrological features, we would recommend to recalibrate the sensor only on the dry days. We added this discussion to the text:*

> ~~With the help of the~~ The methods presented here ~~, we were able~~ can support efforts to identify those residuals to a much higher precision. In order to properly quantify the excess water storages in future studies, we would recommend to calibrate the CRNS signal only in periods when the site had not been exposed to rain events for a few consecutive days. In the case of the *Wüstebach* site (Fig. 9), this would lift the deviations of the CRNS signal (light blue) up from below the averaged soil moisture (blue) and would thus properly pronounce the added water to the system.

1.2. Major concern #2

RC: ***It is very confusing when the authors mentioned "theoretical line" "conventional" "revised", "equal". To my understanding, the "theoretical line" refers to the Equation (1), and the N0 is determined by using the standard sampling scheme as defined by Hydroinnova-soil (e.g. Desilets et al. 2010). For the standard sampling scheme, the equal weighting is deployed for horizontal and vertical samples. Then, how to distinguish this "theoretical line" with "equal" approach? Please help to make all implications for different approaches more explicitly.***

AR: *The procedure we followed is actually less complicated. The theoretical line – the relation between neutrons and water equivalent – is independent of the weighting strategy. Desilets et al. (2010) determined the relation using neutron physics simulations and validated with exprimental data at mostly homogeneous sites. Thus, this relation can be considered as a fundamental relationship of the CRNS method. We found that it can be confirmed at best by observations if the revised weighting approach is used. We added a schematic and flow chart of the procedure to make clear that the N_0 parameter of the theoretical relation can be calibrated using average soil moisture obtained by either the equal, conventional, or revised approach (see Fig. 2 in this very document).*

RC: ***It is recommended to use same symbols for same physical quantities, while using different sub/super-scripts to represent different weighting schemes. Please see my detailed comments in the attached PDF.***

AR: *See the specific comments below.*

RC: ***It is also recommended to list all different methods into a summary table, which will help readers to understand the topic easier than the currently presented.***

AR: *Thank you for this suggestion. In this work we applied three methods: equal weighting, conventional weighting, and revised weighting. These are explained linearly in the methods section. We think that a summary of these methods in a table is redundant and would not add substantial insights about the procedure. Instead, we followed your advice to illustrate more clearly what the methods are and how the weighting procedure works. Therefore we added a single-column figure, including a schmetic of the CRNS footprint, the point samples, and a flow chart of the weighting procedure (see Fig. 2 in this very document). The figure is referenced in the first sentence of section 2.3:*

> The following procedure is recommended to generate a weighted average of point measurements that can be compared with the CRNS product (see illustration in Fig. 3).

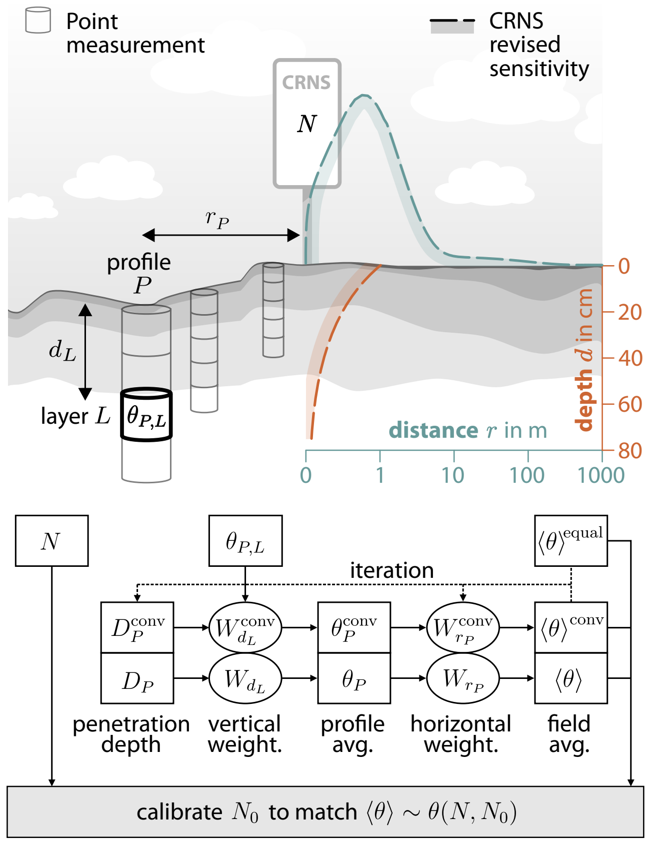

Figure 2: (added to paper as Fig. 3) Top: Schematic of the environment around the Cosmic-Ray Neutron Sensor (CRNS) including point measurements (e.g., soil samples) of water equivalent θ to calibrate or validate the sensor. The revised sensitivity functions W_r (teal) and W_d (brown) are indicated, the omitted y or x axes express their sensitivity at arbitrary scale, respectively. Bottom: The measured variables are used in the weighting procedure (section 2.3), starting with an initial estimate of field-average water content. Three approaches, using the *equal*, the *conventional* (conv), and the *revised* weighting function are compared in this study. The resulting weighted-average water equivalent $\langle\theta\rangle$ is then used to calibrate against or validate with the CRNS product (eq. 1). Calibration of the parameter N_0 is performed towards optimization of four performance measures (see section 2.5).

1.3. P3 L27 eq. 1

RC: ***what is N should be explained as well.***

AR: *N denotes the corrected neutron count rate as explained in the sentence before:*

> To convert the corrected neutron count rate N to gravimetric soil water equivalent, θ, Desilets et al. (2010) suggested the following theoretical relation:
>
> $$\theta(N, N_0) = \frac{0.0808}{N/N_0 - 0.372} - 0.115\,, \quad \text{(Desilets et al. 2010)}$$

1.4. Fig 1 caption: "underestimates"

RC: ***it would be better to add here a constrain, "when compared to the revised vertical weighting function", for this statement.***

AR: *Thank you, we clarified the statement:*

> Normalized vertical weighting functions (eqs. 1 and 2) based on 12 sample points. ~~show that the~~The conventional, linear approach underestimates the relative contribution from shallow water when compared to the revised, exponential function, and neglects contributions from depths beyond ~~z^* (= 23~~ $D^{\text{conv}} \equiv z^*$ (= 23 cm in this example).

1.5. Fig. 1 annotations

RC: ***move title to y axis***

AR: *We like to keep the y-axis label on top of the plot to improve readability (non-rotated text, and larger figure size due to a smaller left margin). We think it is further obvious that the label at the top refers to the y-axis, since a plot by definition shows $y = f(x)$, such that the title of a plot is equivalent with the title of y.*

RC: ***please delete 'revised' here, while use 'rev' as superscript.***

AR: *We do like to keep written text in plot legends as the figure is then easier accessible by readers, compared to only showing a symbol. We further like to keep the convention of using symbols W, D, $\langle\theta\rangle$ for the revised approach, and W^{conv}, D^{conv}, $\langle\theta\rangle^{conv}$ for the conventional approach. Putting ^{rev} on top of the symbols would further complicate the presentation in this mansucript. The revised approach is the main protagonist in this work, and references to these symbols occur many times all over the paper, partly adding further symbol annotations (e.g., W_r^*).*

RC: ***how this ["equal, 1"] is determined? not described in the context, not in the caption as well.***

AR: *Thank you, we added description to the caption that we normalized the weights of 12 sample points (then, the equal weights give 1/12 for each point). Nevertheless, this is an exemplary and arbitrary value.*

> Normalized vertical weighting functions (eqs. 1 and 2) based on 12 sample points.

RC: ***use cm^3 cm^-3 as unit. Please check it throughout the whole manuscript.***

AR: *We do not see urgent needs to use cm^3 cm^{-3}, mm^3 mm^{-3}, or any other equivalent unit. We think volumetric percent (%) is a much simpler unit and integer numbers are easier accessible for readers than decimal fractions. However, in this revision we agreed to follow ISO80000 standards, by removing the indices from units like $\%_{vol}$, and instead added indices to the variables, e.g. $\theta_v = 3\,\%$. We changed this throughout the document. We also changed axes labels for soil moisture in all figures to "Vol. soil moisture θ_v in %". To clarify the convention, we added a note to the methods section:*

> Quantities related to absolute water equivalent are either given in units of gravimetric percent (θ_g in $\% \equiv 100 \cdot \mathrm{g/g}$) or volumetric percent (θ_v in $\% \equiv 100 \cdot \mathrm{cm^3/cm^3}$). If no indices v or g are indicated and units are not mentioned in the context, this work uses default units of volumetric percent.

RC: ***use different color for different r, you can also reduce the number of curves by considering only three different distances, e.g. r=50m 150m, 300m . . .***

AR: *We think that different colors or line styles would not be easily visible in this case. As the line for $r = 0$ is specified, the rank/order of the other lines is self-explanatory, since the curves diverge more and more from $r = 0$ for increasing $r = 50, 100, 150, \ldots$.*

RC: ***Please use a common symbol to denote penetration depth***

AR: *Thank you, we changed to symbol D for penetration depth all over the manuscript. The fact that we still mention z^* and D_{86} has historical reasons, since these symbols are already well known to denote penetration depth in previous literature.*

$$W_d^{\mathrm{conv}} = \begin{cases} 1 - d/\cancel{z^*}D^{\mathrm{conv}}\,, & d \le \cancel{z^*}D^{\mathrm{conv}} \\ 0\,, & d > \cancel{z^*}D^{\mathrm{conv}} \end{cases} \tag{1}$$

$$\text{penetration depth:} \quad D^{\mathrm{conv}} \equiv z^*(\theta) = \frac{5.8}{\theta + 0.0829}\,, \quad \text{in cm, see Franz et al. (2012b).}$$

> The two major shortcomings of this function are (1) that it assumes similar penetration depths of detected neutrons for all distances r from the sensor (see Fig. 1a), and (2) that it neglects any contribution of soil water below a certain cutoff ~~depth z^*~~ penetration depth D^{conv} (see Fig. 1b).

$$W_d(r, \theta, p, H_{\mathrm{veg}}) = e^{-2d/\cancel{D_{86}}D}\,, \qquad \text{penetration depth:} \quad \cancel{D_{86}}D \equiv D_{86}(r^*, \theta, \varrho_{\mathrm{bulk}})\,, \quad \text{in cm, see Appendix A.} \tag{2}$$

> where ~~D_{86}~~ D denotes the effective penetration depth, defined as the depth within which 86 % of neutrons probed the soil

1.6. P5 eq. 2

RC: ***what is denoted by 'd'? geometry depth? Please put it in the text.***

AR: *The variable d denotes the vertical depth of soil moisture measurements as explained in the paragraph before:*

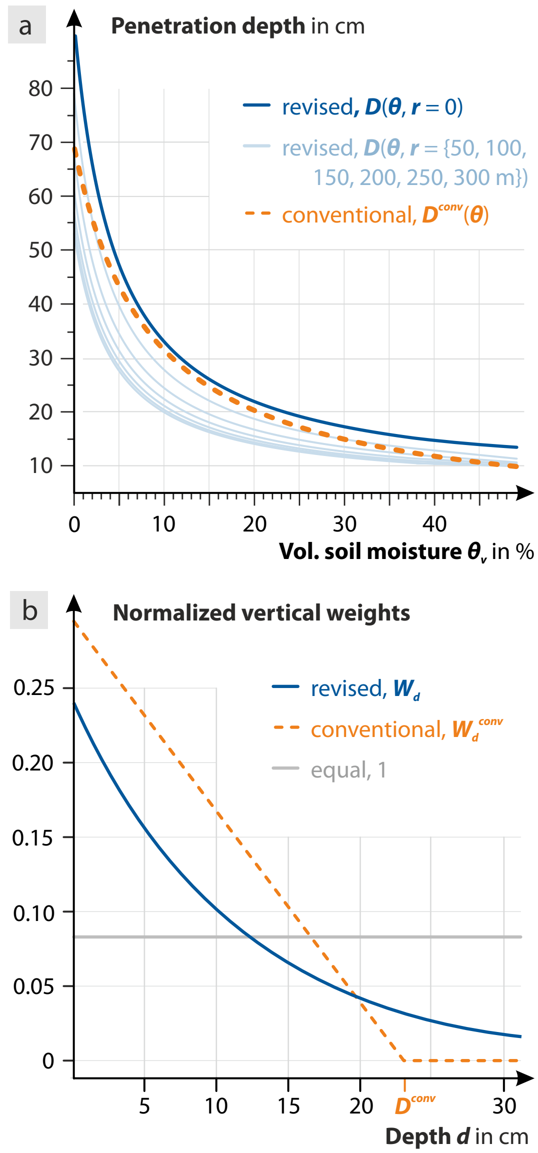

Figure 3: (Updated Fig. 1 in the paper) (a) A comparison between the *revised* and the *conventional* penetration depths, ~~$D_{86}(\theta, r, \varrho_{\rm bulk} = 1.4\,{\rm g/m^3})$~~ $D(\theta, r, \varrho_{\rm bulk} = 1.4\,{\rm g/m^3})$ and ~~$z^*(\theta)$~~ $D^{\rm conv}(\theta)$, respectively. On average, both approaches follow an almost similar shape, however the conventional formulation is independent of distance r and soil bulk density $\varrho_{\rm bulk}$. (b) Normalized vertical weighting functions (eqs. 1 and 2) based on 12 sample points. ~~show that the~~The conventional, linear approach underestimates the relative contribution from shallow water when compared to the revised, exponential function, and neglects contributions from depths beyond ~~z^* (= 23~~ $D^{\rm conv} \equiv z^*$ (= 23 cm in this example).

Therefore, independent soil moisture measurements taken at different depths, d, need to be weighted differently in order to account for the underlying physical processes. To show the consequences of neglecting this step in post-processing, we have compared the *equal* average of soil samples with alternative weighting approaches. The *conventional* vertical weighting, W_d^{conv}, is performed using a linear relation from Franz et al. (2012b), which was based on Monte-Carlo simulations from Zreda et al. (2008) and became widely accepted in most previous studies.

1.7. P5 L5 pdf annotation

AR: *Thank you, we changed as suggested:*

The two major shortcomings of this function are (1) that it assumes similar penetration depths of detected neutrons for all distances r from the sensor (see Fig. 1a), and (2) that it neglects any contribution of soil water below a certain cutoff ~~depth z^*~~ penetration depth D^{conv} (see Fig. 1b).

1.8. Fig 2 caption "underestimates"

RC: ***again, need to be constrained with "when compared to the revised horizontal weighting function"***

AR: *Thank you, we changed as sugegsted:*

The *conventional* approach is insensitive to air and soil water content and highly underestimates the contribution of nearby areas ($r < 10$ m) when compared to the revised function.

1.9. P7 L5 "while estimations of soil bulk density exist for every profile (or even every sample)."

RC: ***Does not 'profile' include 'every sample'?***

AR: *In general, a soil profile could have multiple samples (in different depths). Thus, bulk density information could be available for each profile (i.e., a vertical average), or even for each sample in the profile (i.e., a vertical distribution of bulk density information).*

1.10. P7 L8 "The procedure to obtain a weighted average of soil water equivalent is described as follows"

RC: ***Please use a flowchart to assist your description here. Please also include equation numbers in the flowchart.***

AR: *Thank you for this suggestion. We included a flow chart as Fig. 3 in the paper, which is a new Fig. 2 in this very document. We did not include equation numbers, because all relevant equations are bundled in sections 2.2–2.3.*

1.11. P7 L11

RC: ***as commented above, please use the same symbol to denote penetration depth, while using different sub/super-scripts to distinguish from revised or conventional.***

AR: *We try to keep the same symbols for the same quantities as defined in the literature in order to avoid confusion* *during the calculation. Then, we used the common symbol D to denote the penetration depths in this work. See comment and corresponding text changes above.*

1.12. P8 L2

RC: ***again, same symbol for vertically weighting function, while using sub/super-script to distinguish.***

AR: *We used the same symbol W_d for the vertical weighting functions, while W_d denotes the revised and W_d^{conv} the conventional approach. The subscript L indexes the soil layer corresponding to the measurement depth d_L, as explained in the sentence before:*

> Vertically average the values $\theta_{P,L}$ over layers L, to obtain a weighted average for each profile P:
>
> $$\theta_P^{\text{conv}} = \text{wt}\big(\theta_{P,L}, W_{d_L}^{\text{conv}}\big)\,, \quad \text{or} \quad \theta_P = \text{wt}\big(\theta_{P,L}, W_{d_L}\big)\,.$$

1.13. P8 L5

RC: ***what is the converge criterion? $|\theta_t - \theta_{t-1}| < 0.0001$?***

AR: *The criterion depends on the accuracy level that is relevant for the individual study. Typical uncertainties for CRNS products are at about 0.02 volumetric percent. The convergence criterion could be half of that. However, we would not like to prescribe this value. We clarified as follows:*

> Use the new $\langle\theta\rangle$ to reiterate through steps 1.–5. until values converge within a user-defined accuracy range (e.g., 1 %).

1.14. P10 L3

RC: ***is this N0 calibrated by using the weighting procedure as described in 2.3***

AR: *Yes, we clarified and illustrated this in the new Fig. 3 (Fig. 2 in this very document, see above).*

1.15. Section 3

RC: ***It is recommended to provide the dominant basic soil texture information, bulk density, porosity, soil organic content for all sites considered here. At least for the <10m footprint of CRNS.***

AR: *Thank you, we moved parts from the old Table 1 into a new Table 1 dedicated only for study site description. Therein we added information about bulk density and other hydrogen pools as suggested.*

1.16. P12 L24-25

RC: ***Please help indicating which site is used for calibration, which site is used for validation.***

AR: *We calibrate the CRNS at every single site. The weighting approach is necessary for both, calibration and validation of the neutron sensor. However, in this work we only calibrate the sensor on the given sampling campaigns data or time series data. In this revision, we have avoided the word "validation" in the context of time series data throughout the text to clarify this. Further clarification is also given in the new Table 1 and the new Fig 3 (Fig 2 in this very document). Clarification was also added to the text in section 3:*

Table 1: (added to paper in section 3) Overview of the investigated study sites, their average bulk densities $\langle\varrho_{\text{bulk}}\rangle$ (in g/cm^3), and average volumetric water equivalent $\langle\theta_{\text{offset}}\rangle$ of additional hydrogen pools (e.g., soil organic carbon, lattice water, root biomass, see eq. 2).

site	period	description	$\langle\varrho_{\text{bulk}}\rangle$	$\langle\theta_{\text{offset}}\rangle$	calibration on
Sheepdrove Organic Farm, UK	2015–2016	grassland with central stripe	1.16	9.0	3 sampling days
Braunschweig, GER (Scheiffele et al., 2015)	05–10/2014	irrigation agriculture	1.49	1.0	3 sampling days
Schäfertal, GER (Martini et al., 2015)	2012–2013	heterogeneous hillslope	1.15	5.2	time series
Grosses Bruch, GER	08–12/2012	pasture grassland, floodplain	1.31	10.0	time series
Wüstebach, GER (Bogena et al., 2013)	04–08/2012	forested river catchment	0.83	6.7	time series
T.W.D.E. Forest, US (Lv et al., 2014)	06–09/2012	complex forest, grass, sage	1.10	4.5	time series

At sites 1–2 the CRNS product is calibrated on datasets from so-called *calibration campaigns*. The term typically refers to one or more days on which soil samples were taken in the field and then analyzed for soil water content in the lab. Sites 3–6 provide *time series* data from soil moisture monitoring networks (e.g., SoilNet, see also Bogena et al. (2010)). These datasets are usually applied to validate the performance of CRNS sensors, however, the present study takes advantage of the continuous recordings in order to calibrate the CRNS probe. Table 1 provides an overview of the site characteristics.

AR: *While the general note related to the definition of "validation" in this work was removed:*

~~**General note:** The experimental data collected in these sites can be classified with the following terminology. *Calibration datasets* typically denote soil samples taken at a single point in time, which are then analyzed for soil water content in the lab and used to find a calibration parameter N_0 for the specific site. Notwithstanding, by repeating the sampling on other days also a validation can be addressed to some extent. *Validation datasets* typically denote time series measurements of a soil moisture monitoring network (SoilNet, Bogena et al. (2010)). They are used to validate the performance of the CRNS sensor, but can be also used to find the calibration parameter N_0 with the help of so-called performance measures. In the present work, we are using the terms *calibration site* and *validation site* only to distinguish between temporally distinct samples and continuous data of independent soil moisture measurements, although in both cases those datasets are used to calibrate the CRNS probe.~~

1.17. Fig 4 caption

RC: ***How this line (or N0) is determined? By following the standard sampling procedure as recommended by Hyroinnova-Soil?***

AR: *The theoretical line (including N_0) is determined by calibrating the CRNS on the weighted-average of point measurements as described in the methods section. We added a schematic and flow chart (new Fig. 3, or Fig 2 in this very document) to clarify this.*

1.18. P13 L9, "reveal a single site-specific calibration curve"

RC: ***what are you referring to here? Do you mean unique?***

AR: *Yes, the parameter N_0 can be calibrated with a single calibration day dataset. This would lead to a single curve, $\theta(N, N_0)$. The three calibration days presented here lead to three distinct curves when the* conventional *weighting approach is used. However, since N_0 is assumed to be a site-specific, static parameter, we would expect all calibration datasets to agree with each other. The* revised *approach brings those three datasets in line, such that a unique parameter N_0 (and thus a single curve) can explain all three sampling campaign datasets. We agree to the reviewer that "unique" would be a better choice here and changed as suggested:*

> By choosing the *revised vertical* weighting approach (green), the calibration points become much better in line with each other and reveal a ~~single~~ unique site-specific calibration curve.

1.19. P13 L10-12

RC: ***Please provide quantitative description on the horizontal weights used.***

AR: *We do not understand what the reviewer means with "quantitative description". The horizontal weights W_r are calculated using equations in the methods section, and illustrated in Fig. 2. We added a new flow chart (new Fig. 3 in the paper, Fig. 2 in this very document) to illustrate how the horizontal weights are applied to a dataset of points.*

1.20. Fig. 5 caption

RC: ***unique? site-specific characteristic theoretical line***

AR: *In accordance to a comment above, we also replaced "single" with "unique" here (see changes below).*

RC: ***It is very confusing which line indicating which data set.***

AR: *We tried to make the connection clear by letting the lines pass through their corresponding ellipse. For further understanding, we changed the figure caption as follows:*

> Sizes of the circles indicate the corresponding uncertainty range of the measurement, while every such measurement corresponds to a calibration curve $\theta(N, N_0)$. The *conventional* weighting approach is not able to provide a ~~single~~ unique theoretical line through the three calibration days and further predicts unrealistic reduction of hydrogen pools during maximum plant height (July) and after harvest (Oct). The revised approach converges the datasets to confirm the accepted neutron theory almost in a single calibration curve within uncertainties (ellipses).

1.21. Figure 6, "CRNS fit to revised"

RC: ***did you mean N0=1172 here?***

AR: *Correct. Each parameter N_0 resulted from the calibration using the given weighting approach. As it is written in the plot legend, the CRNS product, $\theta(N, N_0)$ was fitted to (i.e., calibrated against) the soil moisture data of the revised weighting approach. We added a new flow chart as a new Fig. 3 (Fig. 2 in this very document) to better clarify how weighting approaches and calibrations are connected.*

1.22. Figure 6, "CRNS fit to revised"

RC: ***It is not clear how the "baseline" in-situ soil moisture observation was calculated here? (by equally weighting?)***

AR: *We hope that the previous comment answered the question.*

1.23. Figure 8 caption

RC: ***how to explain those curves lower than the revised?***

AR: *See general comment above (major concern #1).*

1.24. Section 4.5, Table 1

RC: ***This is too brief to be a sub-section.***

AR: *Agreed. We moved the content of this section to the conclusions:*

> ~~**4.5 Summary of the analyzed research sites**~~
> ~~The experimental sites used in this study and the corresponding gain for environmental and hydrological research is summarized in Table 1.~~

> The ~~main conclusions are summarizes as follows~~improvements of the CRNS performances for each site are summarized in Table 2, including a note that highlights specific features of the analyses. The study has led to the following conclusions:

AR: *We further moved descriptive parts of Table 1 to an additional table (now Table 1) in section 3. Table 1 (now Table 2) was moved also to the conclusions, site descriptions have been removed, and the results of the performance gain for each site have been added.*

We hope that the additional table and the summary table of the results accomodates the wishes of the reviewers to add more overview elements and structural clarity to the paper.

1.25. P22 L22

RC: ***This means the standard sampling scheme, using equal weighting for vertical and horizontal?***

AR: *Yes. Even both, the equal and conventional approach led to less performance. We rephrased to clarify this:*

> The *revised* averaging of observed point data improved the performance measures ~~when compared to the CRNS water equivalent, $\theta(N)$,~~ of the CRNS product for all investigated sites when compared with the *equal* and *conventional* approach.

1.26. P22 L24 "theoretical line"

RC: ***equal weighting for horizontal and vertical?***

AR: *We hope it is clear from the other comments above that the theoretical line – the relation between neutrons and*

Table 2: (Updated Table 1 (now Table 2)) Summary of the CRNS performances achieved by changing from the *conventional* to the *revised* weighting approach. $RMSE_v$ is in units of volumetric %. CRNS data has been calibrated on three sampling days (sites 1–2), or on time series of a soil moisture monitoring network (sites 3–6). The revised weighting approach improved the performance at all sites, and helped to identify additional hydrological features.

site	KGE	NSE	RMSE	correlation	note
Sheepdrove Organic Farm, UK			$5.3 \to 1.4$		bias from grass stripe
Braunschweig, GER			$1.2 \to 0.6$		data became more consistent
Schäfertal, GER	$0.88 \to 0.93$	$0.81 \to 0.87$	$4.0 \to 3.3$	$0.92 \to 0.94$	incomplete SoilNet coverage
Grosses Bruch, GER	$0.02 \to 0.48$	$-0.71 \to 0.11$	$3.5 \to 2.3$	$0.80 \to 0.78$	remote ponding
Wüstebach, GER	$0.65 \to 0.69$	$0.41 \to 0.65$	$6.7 \to 5.1$	$0.80 \to 0.81$	revealed excess water storages
T.W.D.E. Forest, US	$0.78 \to 0.91$	$0.72 \to 0.82$	$1.3 \to 0.8$	$0.88 \to 0.92$	areal weighting of 4 clusters

water equivalent – is independent of the weighting strategy. Desilets et al. (2010) determined it using neutron physics simulations and exprimental data at mostly homogeneous sites. Thus, this relation can be considered as a fundamental relationship of the CRNS method, which can be confirmed at best by observations if the revised weighting approach is used.

1.27. P23 L11-13

RC: ***This needs to be reconciled. See my comments on Figure 8.***

AR: *Thank you, we rephrased the statement. We further added clarifications to the text as mentioned in the comments above.*

~~The~~ When CRNS is combined with independent soil moisture measurements, the revised weighting approach has the potential to reveal ~~otherwise invisible hydrological features , like~~ hydrological features that were otherwise lost in the signal by overcalibration. The approach improved the accuracy by which the CRNS probe was able to sense total changes of water storages other than soil moisture, e.g., from water in the biomass or litter layer, interception water~~storage~~, groundwater rise, as well as ponding in remote or local areas.

2. Reviewer #2

2.1. Main concern #1

RC: ***The recommended calibration scheme calls for about half of the calibration samples to be taken inside the 20 m radius. This will over-characterize the small area (<20 m) in comparison to under-characterized wider field (>20 m). I would like to see some discussion of how these over- and under-characterization will affect the calibration.***

AR: *Thank you for pointing this out, which is indeed a big challenge related to sampling strategies and interpolation in the CRNS footprint. The recommendation that we aimed to give is that samples should be taken according to the local conditions and correclation lengths. Thus, a large amount of samples below 20 m only makes sense if the correlation length of soil moisture is much smaller than 20 m. In addition, our argumentation is that point measurements should have highest accuracy in regions where the CRNS is most sensitive. Many samples within the first meters will thus lead to a more accurate average for that area and will contribute to more accurate sensor calibration. Consequently, taking less samples far afield would be justified by the very low sensitivity to those regions. Measurement uncertainties can be higher at large distances, as the influence on the weighted average would be marginal.*

> In general, it can be ~~strongly~~ recommended to select ~~about half~~ a significant portion of available sampling points within the nearest 25 m, since ~~the~~ 30–50 % of detected neutrons typically originated in that area. The conventional sampling scheme from Franz et al. (2012a) does not account for ~~40–50 % of detected neutrons that originated in that area .~~ this contribution, which is particularly relevant if local correlation lengths of soil moisture can be below 20–30 m. The number of samples in an area should also represent its areal contribution to the neutron signal, in order to reduce measurement uncertainty in areas where the CRNS probe is most sensitive. This argumentation justifies a lower amount of samples in regions far afield.

A universal quantification of the effect on the calibration cannot be provided, because it highly depends on local structures and conditions, as has been argued in the text. Instead, a dedicated study would be necessary to quantify the effect of the number of sampling points in different distances from the probe. We will further soften and clarify the statements in the revision.

2.2. Main concern #2

RC: ***The concern about time series stems from the fact that distributed sensor data typically do not include the top 10 cm of soil. And because the top 10 cm is critical to cosmic-ray signal, particularly after a precipitation event, these data are not comparable to cosmic-ray derived soil moisture. Placing point sensors close to the surface (0-2 cm) will change your comparison between point sensors and cosmic-ray data. If these shallow depth data are not available, your conclusion regarding the comparison, and by extension regarding the calibration, suffer diminished credibility. I would like to see an assessment of the effect of the lack of near-surface point data on the comparison. Perhaps you can develop an assessment that could add to the uncertainty of the point data, and this will affect the comparison with cosmic-ray data.***

AR: *This is a good point that indeed influences the calibration/validation performance with CRNS. One of our coauthors, Ling Lv, investigated exactly this issue at their site in the T.W.E. Daniel Forest. They burried the shallowest sensors at 10 cm depth and extrapolated data towards 0–9cm using hydrophysical simulations* *with the soil hydrologic model* `HYDRUS`*. This approach significantly improved the comparison of the time series to CRNS. In our study, we made use of these interpolated and extrapolated data. We added this detail to the site description:*

> For each dominant vegetation type, three plots and three subplots within each plot were randomly chosen~~, and TDT sensors were installed~~. Time series data was evaluated from TDT sensors at 10 cm, 25 cm, and 50 cm ~~depth in each subplot (Lv et al., 2014).~~and interpolated up to the surface using hydrophysical simulations (Lv et al., 2014).

At the other sites with time series data, the shallowest soil moisture sensors were burried at 5 cm, while their electromagnetic sensitivity covers distances of up to 5 cm. We are thus assuming that the measurement is representative for the first layer of the soil. However, we fully agree that surface water at 0–1 cm is probably not tracked by conventional sensors. This is why we argue that deviations of CRNS and soil moisture sensors are often related to intercepted or ponded surface water, especially during and shortly after rain events.

In fact, this issue and the detailed assessment suggested by the reviewer is currently part of a separate research project (in which some of the authors are involved), which aims to quantify the uncertainties related to such lacks of information in soil profiles.

2.3. P1 L1

RC: ***"hydrologists" would do***

AR: *Thank you, we changed:*

> In the last years the method of cosmic-ray neutron sensing (CRNS) has gained popularity among ~~soil~~ hydrologists, physicists, and land-surface modelers.

2.4. P1 L4

RC: ***"neutron density" or "neutron intensity", but not "neutrons"***

AR: *Thank you, we changed:*

> However, the signal still may contain unidentified features of hydrological processes, and many calibration datasets are often required in order to find reliable relations between ~~neutrons~~ neutron intensity and water dynamics.

2.5. P1 L8

RC: ***Clarify the meaning. Is it two calibrations per time series (or per location)?***

AR: *Thank you, we clarified:*

> The approach is extensively tested ~~with two calibration and four~~ at six distinct monitoring sites: two sites with multiple calibration datasets and four sites with continuous time series datasets~~from a variety of sites and conditions~~.

2.6. P1 L9

RC: *clarify the meaning. or drop "robustly"*

AR: *We mean that the improvement is robust across different sites, conditions, and performance measures. However, we agree to drop the word as it would require further explanations which is too detailed for an abstract.*

> In all cases, the revised averaging method ~~robustly~~ improved the performance of the CRNS ~~product and even~~ products. The revised approach further helped to reveal ~~otherwise~~ hidden hydrological processes which otherwise remained unexplained in the data or were lost in the process of overcalibration.

2.7. P1 L10

RC: *what is CRNS product? define it*

AR: *We mentioned in line 2: "The sensor provides continuous soil moisture data", we thus feel that the term "CRNS product" is adequately exlained and would thus suggest not to change the wording (see above).*

2.8. P1 L10

RC: *again - clarify the meaning; is it some signal that remained unexplained before, and which you understood and explained here?*

AR: *Thank you, we clarified it in the text (see above).*

2.9. P1 L10

RC: *unclear from the previous sentences what the "presented approach" is; state clearly what the approach is and then expand on it by providing some insights.*

AR: *Thank you, we clarified that we mean the weighting approach and how it differs from the conventional function:*

> The presented weighting approach increases the overall accuracy of CRNS products and will have impact on all their applications in agriculture, hydrology, and modeling.

and added further explanations to a previous sentence in the abstract:

> In this work, a revised sensitivity function is used to calculate weighted averages of point data. The function is different from the simple exponential convention by the extraordinary sensitivity to the first few meters around the probe, and by dependencies on air pressure, air humidity, soil moisture and vegetation..

2.10. P2 L2: 10 to 10^4 m

RC: *why this range? it doesn't follow from previous text; explain, for example by saying that this is the scale at which many relevant hydrological processes operate at the land surface (and give some supporting*

references)

AR: *Thank you, we changed the wording to make clear that we refer to the models cited in the preceding sentence, which operate at that scale.*

> Consequently, ~~there is~~ the corresponding models raise a huge demand for accurate estimations of root-zone soil moisture at scales from 10 to 10^4 m.

2.11. P2 L3 pdf annotations

AR: *Thank you, we adapted the text accordingly:*

> ~~Cosmic-ray neutron sensors~~ Cosmic-Ray Neutron Sensing (CRNS) ~~are~~ is one of the most promising ~~technologies~~ methods for root-zone soil moisture monitoring at the field scale. These instruments are able to continuously measure soil water content averaged over several hectares and up to half a meter depth

2.12. P2 L5 pdf annotations

AR: *Thank you, we changed as suggested:*

> They are one of the few candidates to close the ~~inconvenient~~ scale gap between point data and remote-sensing products

2.13. P2 L16 pdf annotations

AR: *Thank you, we changed as suggested:*

> Despite the unambiguous improvements obtained by corrections and realibration approaches, still some features in many datasets could not be explained ~~by the given theory~~ using current knowledge and consequently seemed to be unrelated to hydrological processes.

2.14. P3 L2 "Their findings extensively describe the footprint volume"

RC: ***remove extensively***

AR: *We do like to keep "extensively" to emphasize the large gain in knowledge about the CRNS footprint for a variety of cases and conditions, while large parts of the footprint characteristics were unknown beforehand.*

2.15. P3 L20

RC: ***I am not sure if this is the right phrase; can intensity be well mixed? maybe say that detectors measure the neutron intensity in the well-mixed pool of neutrons above the land surface"***

AR: *Thank you, we rephrased as suggested:*

> Stationary cosmic-ray neutron sensors (CRNS) are particle detectors that ~~monitor the well-mixed~~ measure the neutron intensity in the ~~air~~ well-mixed pool of neutrons above the land surface

2.16. P3 L23-24 pdf annotations

RC: ***give references; also, air humidity falls under "hydrogen atoms in the footprint"; please, clarify.***

AR: *Thank you, we changed:*

> The neutron signal is predominantly sensitive to the number of hydrogen atoms in the ~~footprint~~soil, but it is also influenced by changes of air pressure, air humidity, and incoming cosmic radiation.

References are already given in the subsequent sentence.

2.17. P3 L24-25 pdf annotations

AR: *Thank you, we changed as suggested:*

> These additional factors can be ~~excluded~~ accounted for by standard correction approaches (Hawdon et al., 2014; Schrön et al., 2015), such that the remaining signal ~~only represents~~ represents only the hydrogen abundance in the soil and biosphere.

2.18. P3 L30-31

RC: ***give supporting references***

AR: *Thank you, we added citations:*

> However, neutrons are sensitive to all kinds of hydrogen in the footprint, hence the variable θ denotes ~~not only~~ the water equivalent of soil moisture, θ_{sm}, ~~but is rather assumed to also include~~ and additional hydrogen pools, θ_{offset}. The latter comprises for example lattice water, θ_{lw} (Dong et al., 2014) , as well as water equivalent from soil organic carbon, θ_{org} ~~, and~~ (Hawdon et al., 2014), biomass, θ_{bio}, and other dynamic contributors, θ_{other}:
>
> $$\theta = \theta_{\mathrm{sm}} + \theta_{\mathrm{offset}}\,, \quad \text{where} \quad \theta_{\mathrm{offset}} = \theta_{\mathrm{lw}} + \theta_{\mathrm{org}} + \theta_{\mathrm{bio}} + \theta_{\mathrm{other}}\,.$$

2.19. P4 L1

RC: ***I wonder if adding another term to represent "other" hydrogen would be good here; "other" could be defined loosely as any known and unknown hydrogen not represented by the previous terms***

AR: *Thank you for the suggestion, we introduced θ_{other} as suggested (see above).*

2.20. P4 L12 pdf annotations

AR: *Thank you, we changed as suggested:*

extension of the analytical fit of the radial sensitivity function W_r to ~~low~~ short distances, $r \leq 0.5\,\mathrm{m}$,

2.21. P4 L19

RC: ***this sentence is self-contradictory; if they are minor, why recommend future investigation? please, rewrite to make a clear point.***

AR: *Thank you, we clarified that those factors are irrelevant for the investigated sites and are thus of minor importance for the conclusions in this work.*

However, those factors are irrelevant for the investigated sites and are thus of minor importance for the conclusions in this ~~manuscript and should be investigated in future studies~~work.

We omitted to suggest further studies here despite the fact that we still think those studies are important in cases where non-conventional sensor locations are used or sensors placed in extraordinarily complex terrain.

2.22. P4 L21-23

RC: ***in homogeneous sites all approaches should give the same result, shouldn't they? please rewrite to make this clear.***

AR: *You are right, this can be considered as a basic principle of weighted averaging. We tried to emphasize that some previous studies applied the equal average only at homogenous sites, such that potential flaws of that approach did not became evident. However, we skipped this part of the sentence as it created confusion and does not add significant information.*

In addition to the *conventional* and the *revised* approach, this work includes the *equal average* weighting strategy (weights equal 1) to compare the performance when the CRNS signal is intuitively treated as a large-area averaging soil moisture product~~, as was done especially at macroscopically homogeneous sites~~.

2.23. P6 L16

RC: ***is this a proportionality sign? if so, put everything to the right of it in brackets to make it clear***

AR: *Yes. We do not see the need for brackets enclosing a full side of an (proportionality) equation.*

$$e^{-r/127} \approx W^{\mathrm{conv}}_{r\leq 300} = \partial_r \mathrm{CFoC}(r) \propto 1 - a_1\, r + a_2\, r^2 - a_3\, r^3 + a_4\, r^4$$

2.24. P7 Figure 2

RC: ***I wonder if the two graphs combined would show the weights better.***

AR: *We decided to split the graph in order to resolve the short-range and the long-range features of this highly non-linear function. If combined into a single graph, the peak below 5 m would become unresolvable for the human eye.*

2.25. P8 L7-8

RC: ***it shouldn't matter whether water contents are gravimetric or volumetric; but male it clear which on you used, and why you would convert to gravimetric, for example.***

AR: *Thank you. We used the volumetric approach as described in the text and rephrased the text to clarify. We agree that the difference may not be significant, however, the gravimetric approach would contradict with the foundation of the weighting functions.*

> It is also possible to calculate gravimetric water content using local bulk densities before step 3, however, ~~URANOS calculations of W_r and W_d have been conducted only for~~ this approach is not recommended since the revised weighting functions have been determined by simulations of homogeneous soil and volumetric water content

2.26. P9 L1-3

RC: ***does this mean that the fancy weighting may not be necessary? or am I misunderstanding this statement?***

AR: *It is not necessary to integrate over volumes around the point samples as long as they are distributed reasonably across the relevant area or soil profile. The corresponding errors have been observed to be far below significance.*

2.27. P7 L7

RC: ***up to this equation the equations were numbered; can you continue numbering them?***

AR: *We think that equation numbers should not be used inflationary. Only equations that are actually referred to in the text exhibit numbers. If it is a convention of HESS to number every equation, we are open to add these numbers.*

2.28. P7 L21

AR: *We changed as suggested:*

> In this manuscript, this estimation is applied exemplarily to the *Schäfertal* site (section 4.2) in order to quantify the errors introduced by ~~non-ideal~~ incomplete coverage.

2.29. P10 L23

RC: ***state N and W or E***

AR: *Thank you, we added units also to other coordinates:*

> The *Sheepdrove Organic Farm* is located at (~~51.528175, -1.467311~~51.528175°N, −1.467311°E, 190 m asl) in the Lambourn catchment in South England.

2.30. P13 L14

RC: ***clarify and expand this***

AR: *Thank you, we added explanations:*

> Furthermore, the experiment clearly shows the importance of a proper positioning of the CRNS probe. If a sensor is dedicated to measure soil moisture in a certain field, it should be ideally placed in that field. CRNS stations at the field border can be biased by different local characteristics, such as land use or soil properties.

2.31. Fig 5 caption

RC: ***it is a single curve, within uncertainties (ellipses)***

AR: *Thank you, we added as suggested:*

> The revised approach converges the datasets to confirm the accepted neutron theory almost in a single calibration curve within uncertainties (ellipses).

2.32. P14 L1-4

RC: ***This is unclear (and contradictory); how do the orange plots show reduced hydrogen? by associating the measurement with a particular line? even so, the difference is small, so perhaps you could discuss this in some more detail.***

AR: *When the same number of neutrons corresponds to different field soil moisture (i.e., different values on the line $\theta(N)$), then other hydrogen sources must have changed that were not statically included in θ. Considering results from the conventional approach at $N \approx 1150$ cph, the calibration line from May indicates 14 % soil moisture, while the line from July indicates 17 %. Therefore, other hydrogen pools must have decreased to achieve the same neutron count for increasing soil moisture. This can lead to the interpretation of reduced biomass from May to July, which contradicts experience that growing maize increases its biomass. We rephrased the text to make this more clear:*

> Using the *conventional* averaging approach (orange), the ~~calibration data~~ three calibration curves in Fig. 5 indicate that similar soil moisture conditions correspond to increasing neutron counts from May to July and October. Since an increase of N always responds to a decrease of water equivalent, this observation could be interpreted as ~~showing a reduced amount of hydrogen pools in the CRNS signal~~ a reduction of biomass during the period of growing maize. ~~This effect seems to be even higher in October than in July.~~

2.33. P14 L13 "water availability"

RC: ***water content, perhabs?***

AR: *We do like to stay with "availability" to emphasize the importance of soil moisture measurements for crop production.*

2.34. Fig 6 visuals

RC: ***equal and CRNS difficult to discern; can you use a different color scheme (or dashed line)?***

AR: *Thank you, we now used the same color for the CRNS observation as was used in the other figures.*

2.35. Fig 6 N_0 values

RC: ***the three N0 values seem close to one another; can you include standard deviations? I bet they will make all three N0 the same, statistically.***

AR: *Although the range of N_0 from 1139 cph to 1172 cph gives the impression of insignificant deviation, this parameter is highly sensitive to the performance of the CRNS soil moisture product. As N_0 is determined from N and θ by calibration (and not by observation or averaging), we cannot provide standard deviations. However, we agree that the Gaussian propagation of uncertainties for N and θ leads to uncertainties of N_0 that are comparable with the mentioned range. The involved uncertainties were addressed implicitely in the text:*

> Therefore, calibration results that resulted in an RMSE~~of $\approx 4\,\%_v$~~$_v$ of $\approx 4\,\%$ (Fig. 6) are not meaningful under wet conditions (where $\Delta\theta(N) \geq 4\,\%_v$), and are still uncertain under dry conditions (where ~~$\Delta\theta(N) \leq 4\,\%_v$~~$\Delta\theta_v(N) \leq 4\,\%$). Consequently, the partial coverage of the CRNS footprint by the irregularly distributed SoilNet hampers the proper evaluation of the CRNS data, and especially of the weighting strategies.

2.36. Fig 6

RC: ***this right panel is misleading: of course the CRNS with revised calibration will show better agreement with revised data; can you make it clear and include a statement or two to discuss this artefact?***

AR: *Thank you, we clarified this in the caption:*

> The ~~*revised*~~average soil moisture using the *conventional* weighting approach (orange) exhibits poor performance against the CRNS signal (not shown). The *revised* approach improves four performance measures of the averaged soil moisture (blue) and the CRNS signal (light blue), although the SoilNet probes are unevenly distributed in the CRNS footprint.

2.37. P15 L5-6

RC: ***this would be true if the remaining (not covered) area has soil moisture that does not show any correlation with soil moisture in the covered area; as this cannot possibly be true, this statement is also untrue.***

AR: *Thank you, we clarified this statement, saying that 36–51 % of the neutron variability does not directly respond to the wetness conditions monitored by the irregularly distributed network, and in most cases the soil moisture of the outer area can be assumed to correlate with the inner area:*

> As a consequence, 36–51 % of the neutron variability ~~cannot be explained~~ does not directly respond to the wetness conditions monitored by the irregularly distributed network. However, in most cases the soil moisture of the outer area can be assumed to correlate with the inner area.

2.38. P15 L16 “dominance”

RC: ***this is not the best word here; perhaps something simple “As the revised approach shows the best accuracy...”***

AR: *Thank you, we changed as suggested:*

> As the ~~dominance of the revised approach~~ revised approach shows the best accuracy in all four statistical measures~~is evident~~, the RMSE$_v$ is still higher than the measurement error of the daily mean ($\cancel{\approx 2\,\%_v}$$\approx 2\,\%$).

2.39. P16 L1-2

RC: ***state that this is speculative***

AR: *Thank you, we changed:*

> This indicates that deviations can be attributed (1) to the insufficient coverage of the SoilNet, and (2) to different processes in different parts of the footprint (~~e.g.~~ speculative examples are: vegetation growth, forest water interception, snow accumulation, evapotranspiration, plowing, etc.).

2.40. P16 L12-13

RC: ***I think this statement must be justified somehow; how about producing “cow water equivalent” (CWE) and using it to correct the data?***

AR: *This procedure would be ideal to investigate the influence of cows. However, no data is available on which days the cows were present at the field and near the CRNS probe. We softened the statement slightly:*

> Additionally, beginning in the mid of September~~a significant amount of cows were~~ , many cows had been present at this site, which ~~is assumed to lead~~ are assumed to have led to large variations of the neutron signal and thus to a non-meaningful expression of correlation-related measures.

In order to justify the potential effect of cow equivalent water, we can at least argue with preliminary results from recent experiments using “Scientist water equivalent”. The following figure (not included in the manuscript) demonstrates that (1) the influence on the neutron signal depends on the distance of the scientists (or cows), which would require advanced cow-tracking technology at our site, and (2) the sensitivity to scientists at distance r confirms the shape of the revised sensitivity function W_r used in this manuscript.

2.41. P15 L20-22

RC: ***so what’s the point of this exercise?***

AR: *Since the CRNS signal is influenced by intercepted and ponded water, calibration against soil moisture is an uncertain venture. Thus, we repeated the calibration only in periods were soil moisture can be assumed to have the main influence on CRNS. This experiment helps to find out whether the performance gain was just an artefact of the dynamics unrelated to soil moisture. We clarified:*

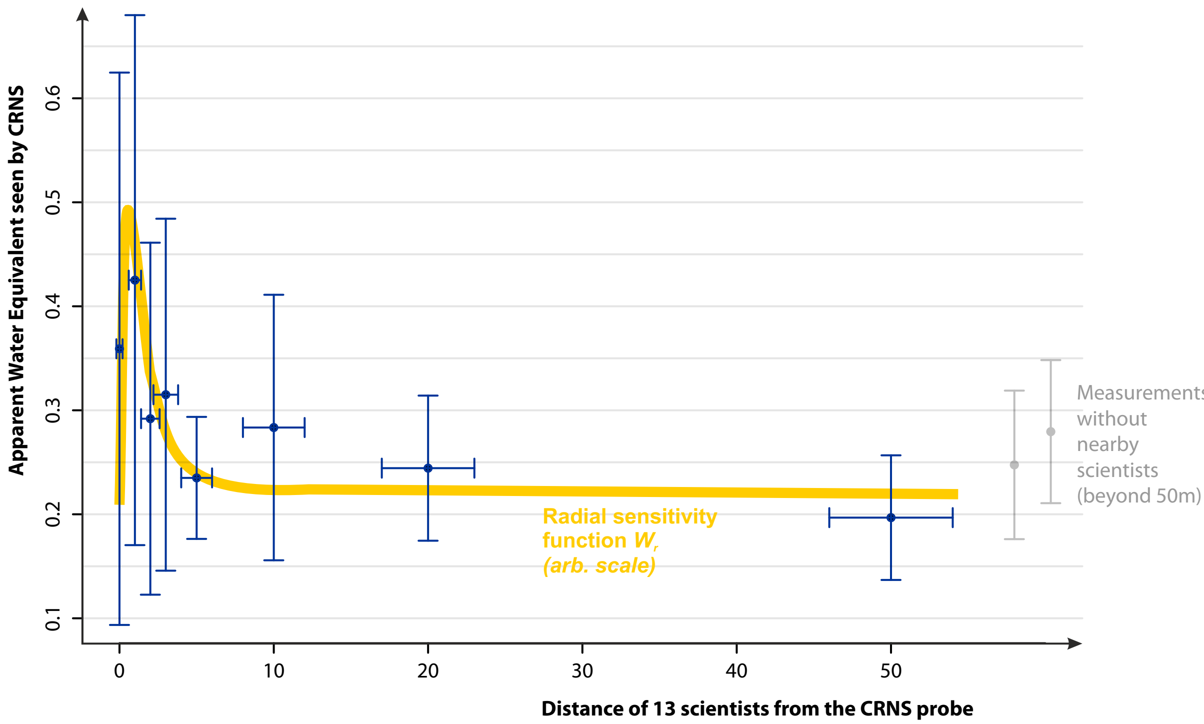

Figure 4: (not in the paper) An experiment to investigate the sensitivity of neutrons to scientists at various distances from the CRNS probe (preliminary). The experiment indicates that also cows could have substantial influence on the neutron count rate (and thus performance) in *Großes Bruch*, where cows are able to approach the CRNS probe as close as 20 centimeters.

> In order to avoid incorrect conclusions from overcalibration of the data during rain events (periods of high interception water), we repeated the same analysis for dry periods only~~, which however resulted in the same conclusions~~. In this case the revised approach again led to highest performance (not shown) and confirmed the robustness of this approach.

2.42. P15 L24-25

RC: ***or moving infiltration front, as you mentioned before; or is this taken care of by the proper weighting?***

AR: *The proper weighting should take care of inhomogeneous soil moisture profiles, although uncertainty may be involved due to the lack of continuous data in the profiles. Detailed investigation of the corresponding uncertainties will be dedicated to another, currently active study.*

2.43. P15 L26

RC: ***so that installation bias can lead to bias in area-average soil moisture based on averaging sensor data; can you elaborate on this possibility as it is as likely as bias in CRNS data.***

AR: *In general this is correct for soil profile monitoring (see general reviewer comments above). However, the statement in the text refers to the difference between soil profile measurements and water storages at or above the surface. Therefore we see no need to change the present formulation:*

> Ponded water in local hollows, trenches, and the litter layer ~~is not visible by~~ are not visible in the soil profiles of the monitoring network, which are typically installed in solid and elevated ground.

2.44. P15 L29 "By comparing CRNS data and point measurements, residual information could be used to identify additional processes like biomass dynamics or rainfall interception"

RC: ***if, and only if, the point measurements are correct AND give true average moisture field; neither is assured.***

AR: *We agree on this comment, however this is a general issue (or assumption) for soil moisture monitoring networks. We thus see no need to add this comment to the argumentation.*

2.45. P13 L10-12

RC: ***so that is the same as giving the average within each land use type to each 1 m2 pixel within this type? what if there is a significant variation within land use type?***

AR: *Yes. Spatial interpolation is always site-specific and should be guided by expert knowledge about local conditions. In our case, mean soil moisture in the four parts had significantly distinct values (as shown by dotted lines in Fig. 9) and variations within the land use types were insignificant (not shown). Since the CRNS represents a large-scale average of soil moisture, we think it is reasonable to smooth small-scale variability of point measurements among an area of comparable characteristics.*

> We therefore grouped the soil moisture information of the four compartments, and weighted each 1 m^2 pixel of the areal contribution map depending on the pixel's distance r to the CRNS probe (see last paragraphs of section 2.3).

2.46. P19 L7 "{25m,75m,200m}"

RC: ***I see this as 25, 75, 175 in Zreda et al. (2012), which suggests rounding to the nearest 25 m; perhaps what Franz et al. (2012) did was to round up the 175 m value to 200? or maybe it is higher because Franz's field site was not at sea level, and thus the footprint was be larger.***

AR: *The latter is true. They roughly adapted the sample radii on the altitude of the individual sites, refering to the dependency of the footprint radius on air pressure as was published by Desilets and Zreda (2013). I guess the authors further rounded the numbers 25, 75, 200, and 175 to the nearest 25 meters for the sake of simplicity. This approach might have had reasons at their individual, almost homogeneous sites, however we think that more accurate sampling and weighting should be performed in any heterogeneous terrain. Which motivated us to write this manuscript.*

2.47. P19 L9

RC: ***These cutoff values seem little too high for exponential decrease with distance according to exp(-distance/127), which you used before. Can you double check it? Is the value of 127 m computed with the increasing area as the distance increases? (It looks like it to me, but please confirm.)***

AR: *Thanks for finding this detail. It is true that the weighting functions account for the increasing area as the distance increases. The cutoff values are indeed too high, as we calculated them by integrating the conventional function beyond 300m. However, as written in section 2.2, Bogena et al. 2013 only defined it for r<300m. We recalculated the cutoff values and adapted the equation. The new values even better match the conventional sampling radii (25,75,175) in the areal centers.*

$$\frac{1}{3}\int_0^\infty W_r^{\text{conv}} \approx \int_0^{59\,48} W_r^{\text{conv}} \approx \int_{59\;\,48}^{186\,142} W_r^{\text{conv}} \approx \int_{186\,142}^{\infty} W_r^{\text{conv}}\,.$$

AR: *Please find also attached a figure that shows the cumulative fraction of neutron count sensitivity over distance. The function from Bogena et al. 2013 (which we used) is only an approximation to the exponential function.*

2.48. P19 L11-12

RC: ***change to "more steep, particularly at short distances to the probe"; away from the probe the shapes of the response functions of Zreda/Desilets and of Koehli are similar.***

please specify that this is total wetness, not only soil moisture, including air humidity and bio-water and other sources.

AR: *Thank you, we changed as suggested:*

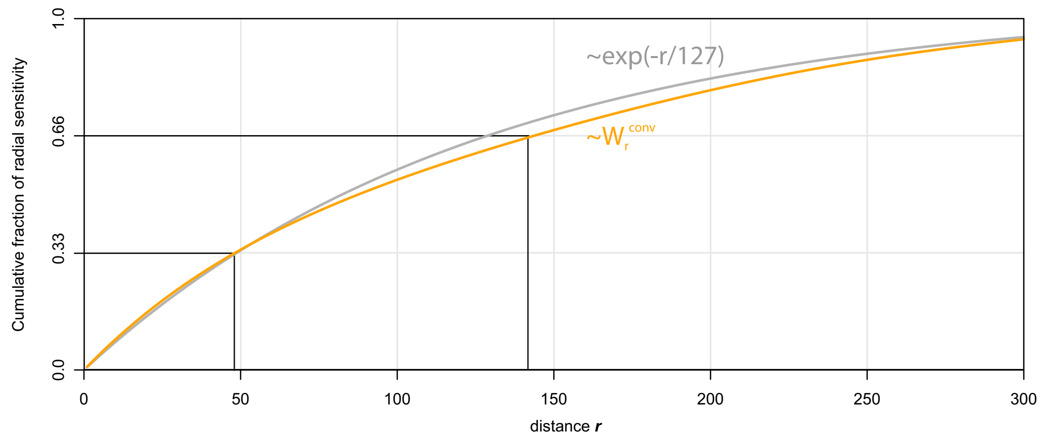

Figure 5: (not in the paper) Cumulative fractions of radial sensitivity, dN/dr, for the conventional weighting functions (exponential, and approximation by Bogena et al. 2013).

> As Köhli et al. (2015) introduced the revised weighting function $W_r(h, \theta)$, the standard sampling scheme has become inappropriate, at least in non-homogeneous terrain, for two reasons: (1) the revised sensitivity is more steep, particularly at short distances to the probe, and (2) ~~dynamically depends on wetness conditions~~depends on the total water equivalent of the surrounding hydrogen pools.

2.49. P19 L13-15

RC: ***this is a very important point; I reviewed a paper in which most of the the calibration points were more than 100 m from the probe, in effect producing questionable (at best) calibration; can you elaborate on this to make it sound stronger/clearer?***

AR: *We dedicated the whole section on this topic, tried to recommend to take more samples within the nearest few meters, and concluded that sampling schemes should depend on local site conditions rather than follow a generel convention. We think we emphasized these results in the conclusions section, particularly in item 5:*

> Data points in the first 0 to 10 m radius and 0 to 20 cm depth around the sensor are most important for calibration and validation purposes. It is thus recommended to reduce the uncertainty of those measurements, e.g., by ~~avoiding flints in the samples, or by~~ increasing the number of samples in that area.

2.50. P19 L18-19

RC: ***not sure if this is the only problem; the calibration problem consists of many subproblems, of which two have to do with spatial variability of soil moisture; one is how representative the data points are; the other is the weights; if you collect half of the calibration points within 25 m, the area enclosed within that 25 m will be well (subjective term) characterized; the other half of the points will be, I guess, beyond the 25 m perimeter; regardless of how far you will go, that area will be less well characterized because you will have fewer points per unit area there than you had in the area between 0 m and 25 m. so the way out***

would be to increase the number of calibration points beyond the 25 m perimeter. How exactly this would work out? I have struggled myself to find out, as it would depend on the degree of heterogeneity of soil moisture. A brief discussion of this issue would make the paper more correct and more relevant.

AR: *Thank you for pointing this out, this is indeed a challenging task shared by all of the evaluated calibration approaches. But, along with all the aspects discussed in the manuscript, the sampling scheme needs to be feasible. As a result, the number of calibration points is practically limited to a number that can be handled within e.g. one day. As a consequence, we think that there is the clear need to optimize the sampling scheme also under the aspect of feasibility. In general we think it is highly advisable to take more measurements closer to the footprint than far afield for the following reasons: (1) the CRNS probe is most sensitive to the first few meters, hence the measurement uncertainty by the soil samples in that area should be as low as possible; (2) the CRNS probe is less sensitive to remote areas, such that uncertainty due to soil sampling has less impact to the calibration result. We changed the statement as follows:*

> In general, it can be ~~strongly~~ recommended to select ~~about half~~ a significant portion of available sampling points within the nearest 25 m, since ~~the~~ 30–50 % of detected neutrons typically originated in that area. The conventional sampling scheme from Franz et al., (2012) does not account for ~~40–50 % of detected neutrons that originated in that area .~~ \DIFadd{this contribution, which is particularly relevant if local correlation lengths of soil moisture can be below 20–30 m. The number of samples in an area should also represent its areal contribution to the neutron signal, in order to reduce measurement uncertainty in areas where the CRNS probe is most sensitive. This argumentation justifies a lower amount of samples in regions far afield.

2.51. P19 L21 "sampling radii R_i of concentric circles could be calculated"

RC: ***actually, wouldn't it be better not to sample on these circles, but within the areas they encompass? it may or may not make much difference, but it would look better (more representative) when mapped. Or perhaps this is, in effect, what you will accomplish by increasing the number of circles from three (as in Zreda 2012) to more than three, as you discuss below.***

AR: *We actually recommend to sample within the area which the circles of radii r_i encompass. It should not be confused with the radii R_i at which the actual sampling takes place. The sentence to which the reviewer refers however deals with the sampling radii R_i. We think this is clear from the text:*

> To give further advice on a reasonable distribution of points for homogeneous terrain, sampling radii R_i of concentric circles could be calculated as follows. First, select a total number of circles n based on prior knowledge about the patterns at the individual site. Since the signal contribution of an area between any radii can be calculated by integrating W_r (compare also Köhli et al., 2015, eq. 1), the n borders of equal areal contribution, r_i, $i \in (1, ..., n)$, can be calculated by solving the integral: ...

2.52. P20 L10

RC: ***equal number will not give equal quality of the average - see my comment two comments above; we would need more sites away from the probe***

AR: *This is true for a sensitivity function that is proportional to $1/r$. The sensitivity functions used here are highly non-linear, such that remote areas have a much less influence on the neutron signal. Given the small range of their contribution, we think it is not required to generate the same measurement quality (i.e., precision) in* *near and remote areas. See also text changes in the next comment.*

2.53. P20 L11

RC: ***I think this is bordering on nonsense - what if the <2 m is affected by the installation of the probe or probes, which is quite likely?***

AR: *If this is the case, the corresponding material moisture should be taken into account (e.g., concrete at $\approx 8\,\%_v$ moisture content). However, most CRNS stations known to the authors do not require significant preparation of the ground, as the detector is often mounted on a thin metallic stick burried in the ground. We further like to emphasize again that the reocmmendations given here have an exemplary status. They cannot be generally applicable, since we argued further in this section that the best sampling scheme always depends on local heterogeneity and site conditions.*

RC: ***Also, if I remember correctly, the maximum sensitivity is NOT at the probe, but some meters away (because of the increasing area), so the <2 m is not justified. Please, rethink and provide an updated figure.***

AR: *The maximum sensitivity is found within the first meter (Fig. 2). We therefore think that it is required to achieve high precision of the soil moisture measurements within this area.*

RC: ***Also, make clear that each of the distances 58, 137 and 240 gets four probes; your descriptions may be understood that all together get 4 locations. (Or simply replace 3x4 with 12).***

AR: *Thank you, we clarified this.*

> ~~Therefore~~Based on this picture, an equal ~~amount~~ number of sampling locations is recommended in each annulus. For example, ~~using the hitherto common amount~~ if it is desired to use the hitherto proposed number of 18 locations ~~, we recommend~~ for humid conditions~~to select 3 locations~~, three could be distributed within 2 m distance, another ~~3~~ three within 17 m, and the remaining ~~3 × 4 locations distributed~~ 12 locations evenly within 58, 137, and 240 m, respectively.

2.54. P20 L16 "compare also Heidbüchel et al., 2016, section 4.2"

RC: ***but Heidbuechel took it straight from Koehli 2015, so it is not independent***

AR: *Correct, but we did not state that the research is independent. In contrast, we mentioned in the Introduction that Heidbüchel et al. 2016 already applied an adaption of the revised weighting functions. Their study serves as a good example here. See P3 L5:*

> ~~Heidbüchel et al. (2015)~~ The revised footprint and spatial sensitivity of the CRNS probe has since been confirmed by many observations. Heidbüchel et al. (2016) were the first to test the impact of the revised ~~spatial sensitivity~~ function on the performance of their calibration data. ~~Encouraged by~~ Schattan et al. (2017) applied the revised weighting approach to average complex snow patterns in an alpine terrain. Encouraged by both of their promising results, the present study has hypothesized that this new theory could enable an improved performance of CRNS calibration and validation campaigns for a huge variety of sites and conditions.

2.55. P20 L20

RC: ***which we almost never have, so this is a moot point; we will have to use guesswork initially to select locations, followed by proper data weighting in postprocessing.***

AR: *We agree, that is true for most applications. We changed the text accordingly:*

> The actual partitioning should rather be guided by expert knowledge about local ~~correlation lengths of spatial soil moisture patterns.~~ patterns, ideally including spatial distributions of soil characteristics and land use. Proper weighting of sampling data in post-processing can be helpful to compensate the lack of such information. Given entirely homogeneous soil, for instance, a single location would do.

2.56. P21 L7

RC: ***I would add the fourth: locations based on land cover type (as you discussed for example in Fig. 9 and elsewhere). If multiple land covers are present, we need to assess each of them in terms of the contribution to neutron count. Or is this covered in your item 3?***

AR: *Thank you, we clarified:*

> Apart from sophisticated optimal sampling designs, three of the most simple sampling strategies are (1) regular grids, (2) random locations, and (3) locations that represent stable ~~soil moisture patterns~~patterns (of soil moisture or land cover).

2.57. P21 L11-13

RC: ***I am not sure what you mean here. Can you rewrite to make clear?***

AR: *Thank you, this paragraph presents another approach to select locations of point measurements in the footprint. The approach accounts for the radial shape of the sensitivity function, however it is independent of dynamic variations. We rewrote the text to clarify:*

> ~~In some cases it could help~~ A simple and pragmatic way to design a reasonable sampling scheme could be to choose sensor locations ~~just in the way it is described by the~~ based on the approximated horizontal sensitivity function W_r^* (Appendix A). ~~Under these conditions~~ As this function does not depend on dynamic changes of surrounding hydrogen pools, an equal average ~~is~~ would be sufficient in post-processing mode.

2.58. P22 L10-12

RC: ***This paragraph is not necessary. It says little and it says it poorly.***

AR: *Thank you, we deleted this sentence.*

~~This study has focused on the theory and application of the averaging approach, while the performance of different interpolation strategies might depend on local soil patterns and deserves a study on its own, for their performance always depend on the local structures and correlation lengths of soil moisture.~~

2.59. P22 L24-26

RC: ***this is the most important point from the practical point of view - it means that we do not need 2+ calibration campaigns, as suggested by Heidbuechel for example, but a single one, properly weighted, will do. This is a great result! You should emphasized it more strongly.***

AR: *Thank you, we tried to emphasize this key point:*

> The results show that unrealistic deviations of multiple calibration datasets from the theoretical line can be removed by applying the revised weighting functions. Thus they support the original hypothesis by Desilets et al. (2010) of a single calibration campaign to capture the local soil moisture dynamics. The approach can thus substantially reduce the calibration efforts for CRNS probes, in contrast to recent findings from Iwema et al. (2015) and Heidbuechel et al. (2016).

2.60. P22 L29

RC: ***remove "as has been presented in section 2.4 and section 4.2".***

AR: *Since information and methods in this manuscript are already rather dense, we do like to keep directives in the conclusion to guide the reader to the relevant parts of the text.*

> Although existing data can be weighted in post-processing mode, missing locations close to the detector as well as insufficient coverage of the CRNS footprint introduce significant uncertainty. It can be quantified with the help of the radial sensitivity functions, as has been presented in sections 2.4 and 4.2.

2.61. P22 L30 "Sampling strategies that are based on concentric rings can only be recommended for homogeneous terrain"

RC: ***but this is not known a priori, so this recommendation is not very useful to a field hydrologist.***

AR: *State-of-the-art research about optimal design and sampling strategies recommends to take into account additional information (proxies) about local structures at the individual sites (Diggle and Lophaven, 2006; Nowak et al., 2010; Bramer et al., 2013; among others). We therefore think it is reasonable to recommend* a priori *site assessment in the footprint of the CRNS probe in order to achieve best results for the sensor calibration.*

2.62. P23 L7 "by avoiding flints in the samples"

RC: ***remove this, or replace with something like "by avoiding stones"; but even this seems too vague or far fetched as the flints or stones will contribute to the neutron signal; increasing the number of samples is the best bet here.***

AR: *Thank you, we changed as suggested:*

> It is thus recommended to reduce the uncertainty of those measurements, e.g., by ~~avoiding flints in the samples, or by~~ increasing the number of samples in that area.

2.63. P22 L11-13

RC: ***so how is ponding otherwise invisible? and, more importantly, would you be able to see these things if you look at the neutron data alone? would you be able to separate interception from ponding? I have doubts, and should you have similar doubts, please spell it out.***

AR: *Thank you for the clarification, we rephrased the text to make this clear. Regarding the invisibility: Ponding and other water storages are invisible for conventional soil moisture sensors. If the CRNS is calibrated against the wrong average, the relevant changes of neutron counts could be interpreted as noise, and identification of other water storages could become impossible.*

> ~~The~~ When CRNS is combined with independent soil moisture measurements, the revised weighting approach has the potential to reveal ~~otherwise invisible hydrological features , like~~ hydrological features that were otherwise lost in the signal by overcalibration. The approach improved the accuracy by which the CRNS probe was able to sense total changes of water storages other than soil moisture, e.g., from water in the biomass or litter layer, interception water~~storage~~, groundwater rise, as well as ponding in remote or local areas.

I agree that the CRNS estimates total storage and by that, it does not distinguish the different features. But we could distinguish them by considering their different temporal dynamics. This was suggested by Desilets et al (2010) and it was used in Baroni and Oswald (2015). In the latter study, seasonal overestimation was attributed to growth biomass. This signal was then removed. The remaining subdaily residuals were analyzed and attributed to water interception and by identifying also a daily cycle.

3. Reviewer #3

3.1. P.3, L.2

RC: ***delete 'how'***

AR: *Thank you, we changed as suggested:*

> This revision has since changed the way ~~how~~ CRNS measurements are interpreted.

3.2. P.3, L.26

RC: ***for improved clarity, insert 'corrected' before neutron count rate.***

AR: *Thank you, we changed as suggested:*

> To convert the corrected neutron count rate N to gravimetric soil water equivalent, θ, Desilets et al. (2010) suggested the following theoretical relation:

3.3. P.8, L.20-25

RC: ***I fail to see the 'regular grid' in Fig.9***

AR: *Thank you for pointing out the lack of visual clarity in Fig. 9. The colored areas therein have a pixel size of 1 m – only visible at high zoom level. We improved the text accordingly:*

> If sample locations were arranged in an interpolated, regular grid (~~see~~ e.g., pixels of size 1 m in Fig. 9), then each pixel should be weighted individually as a point such that the integrals above can be simplified.

Caption of Fig. 9:

> The areal coverage was then averaged (dashed line) pixel by pixel (1 m resolution) with the revised weighting approach, leading to the best performance against the CRNS signal.

3.4. P.8, L.20-25

RC: ***I do not follow the derivation of the sector weighting, and as written, it is mathematically incorrect (does it not give the result of Wr/12Pi?***

AR: *We admit that the calculation was a bit vague and overcomplicated, so we improved the text as follows:*

~~For example in a polar grid with 6 sectors, each sector at distance r is to be weighted with $\int_{\text{sector}} W_r = W_r/2\pi r \cdot (r/6) = W_r/(6\pi)$. In a rectangular grid of grid~~ While an infinitesimal point at distance r has the weight $W_r/(2\pi\, r)$, a regular pixel of size s ~~, the number of pixels per ring,~~ at that distance weighs $W_r/(2\pi\, r) \cdot s \propto W_r/r$. For all radially symmetric sampling schemes, where each point measurement represents one of n ~~,~~ circular sectors, the sector at distance r ~~is $n(r, s) = r/s$, such that the weight for each pixel is to be $\int_{\text{pixel}} W_r = W_r/(2\pi r) \cdot (r/n) \propto W_r/r$~~ has the size (arc length) of $2\pi\, r/n$, and thus contributes the weight $W_r/(2\pi\, r) \cdot (2\pi\, r)/n \propto W_r$.

3.5. P.8 L.24

RC: ***grid size s is presumably a dimension - is that the length of the side one pixel?***

AR: *Correct, please see above.*

3.6. P.8, end of L.25

RC: ***presumably should be Wr/n?***

AR: *Correct, while $n = r/s$. This leads to the said result, $\propto W_r/r$.*

4. Proactive changes by the author

AR: *The following changes were initiated by the authors alone.*

4.1. Titel

Following the requests by Reviewer #1 to simplify the presentation of the paper, we have also reconsidered the title. We now came to the conclusion that the last phrase "Theory and Evidence" is both unattractive and redundant, while "evidence" does not properly reflect the accurate quantification of performance measures in our work. We thus like to change the title as follows:

> Improving Calibration and Validation of Cosmic-Ray Neutron Sensors in the Light of Spatial Sensitivity ~~– Theory and Evidence~~

4.2. Column layout for figures

Because the final HESS format will be in two-column style, we rearranged the subfigures in Figs. 1, 3 (now 4), and 11 (now 12), such that they would fit into one of the two columns (i.e., no column spanning anymore for these figures).

4.3. P24 eq. A1

We do like to correct a typo in the definition of r^* (eq. A1). In the submitted manuscript, the correction factors for air pressure and vegetation were multiplied by r. However it is indicated in the text and in the figure that the relation should be inverse. The typo has been slipped in to eq. A1 and to the supplements, while procedures of the performed analysis remained unaffected. We changed the manuscript and the supplementaries accordingly.

$$\text{where} \quad r^*(r, p, H_{\text{veg}}, \theta) = r \cancel{\cdot} / F_p \cancel{\cdot} / F_{\text{veg}}(H_{\text{veg}}, \theta)\,. \quad \text{(A1)}$$

4.4. P3 L5-7

A research article from Schattan et al. (2017) has been recently published in WRR. The authors confirmed the revised footprint radius as well as the revised weighting function in a complex alpine terrain. We would like to include their findings as an additional motivation in our introduction:

> ~~Heidbüchel et al. (2015)~~ The revised footprint and spatial sensitivity of the CRNS probe has since been confirmed by many observations. Heidbüchel et al. (2016) were the first to test the impact of the revised ~~spatial sensitivity~~ function on the performance of their calibration data. ~~Encouraged by~~ Schattan et al. (2017) applied the revised weighting approach to average complex snow patterns in an alpine terrain. Encouraged by both of their promising results, the present study has hypothesized that this new theory could enable an improved performance of CRNS calibration and validation campaigns for a huge variety of sites and conditions.

4.5. P3 L22

As Köhli et al. (2015) have shown, 86 % of detected neutrons can come from distances of up to 240 m at sea level and under dry conditions. We like to change the description in the methods section slightly to emphasize that there is ofcourse a contribution of neutrons beyond 240 m, especially at high altitude (low air pressure).

> Due to the low interaction probability of neutrons with air molecules, the measured particles can travel distances of ~~up to~~ more than 240 m from the soil to the detector