Response to Interactive comment by H. Bogena (Referee)

Comments from referee are printed in black. Authors’ responses are printed in red.

Résumé

This study concerns the application of the cosmic-ray neutron method to monitor soil moisture in a mixed forest in the lowlands of north-eastern Germany. The authors tested several calibration procedures using soil samples taken 10 times within one year inside the footprint of the sensor. A two-point calibration is assumed to be adequate to correctly define the shape of the N0-calibration function with adjusted parameters when calibration points were taken during both dry and wet conditions covering at least 50% of the total range of soil moisture.

This paper is an interesting presentation of a forest application of the cosmic-ray neutron method giving some new insights into the calibration process. It is also very well written and fits well to the scope of HESS. However, some methodological improvements need to be undertaken as outlined in my comments. In addition, the results should be better discussed in the light of existing publications on the calibration of the cosmic-ray neutron probe.

We want to thank Heye Bogena for a detailed and thoughtful review that helped in significantly improving our manuscript.

General comments

The calibration results might have been affected by the unfavourable locations of the sampling locations. As demonstrated by Köhli et al. (2015), the highest contribution comes from the first 10 m radius, whereas the nearest sampling locations are still 25 m away from the CRNS probe. Please add a discussion on how thus soil moisture differences between the close-up area and the sampling locations might have affected the calibration results.

Fortunately, the conditions within the closest 30 m around the CRS are quite homogenous at our field site since the sensor is located within a pure beech stand. So we expect a similar range of soil moisture conditions at our sampling locations at a distance of 25 m away as compared to those at locations closer to the sensor. Thus, the near-field sampling design should not influence the calibration results. Our aim was to make the sensor work for our purposes in our environment. However, in the future we have to think about what it really is that we are measuring. Is the CRS a measuring device that can only give us reliable information on soil moisture within a radius of 10 m? Maybe the change in the shape of the calibration function that we found is also a result of trying to counteract/compensate for spatial heterogeneity. We added some of this discussion to the manuscript (in section 4.2).

The vegetation within the footprint of the CRS probe is quite heterogeneous (please add a table of the landuse contributions), which complicates the spatial averaging of soil moisture. For instance, as shown by the authors, the coniferous sites are consistently dryer compared to the areas covered by beech. Thus, the limited number of sampling locations could be an additional reason for the differences in the CRS calibrations.

This is true. We state in the manuscript that this variability within the footprint is a possible cause for the differences in our calibration results. We also tested whether an area-weighting function based on the different tree stands significantly changes our results. We found that it did not make a difference. We will add a table of land use as recommended.

Iwema et al. (2015) already showed that using only one calibration data can lead to large uncertainties especially in humid regions with large hydrogen pools. They showed that the best trade-off between number of calibration dates and calibration accuracy can be achieved by using 6 calibration dates. Please discuss your results in the light of the results of this recent publication.

As far as we understand, Iwema et al. state that using more than 6 random sampling dates does not further improve the calibration results. They also state that using sampling dates with dissimilar moisture conditions (‘appropriate soil wetness conditions’) can reduce the required number of sampling dates. We found that using only 2 sampling dates with a large enough difference in soil water content suffices to achieve a good calibration. We added a discussion of Iwema et al. (2015) to the section 5.1 (lines 687-695).

This study also considers the vegetation correction developed by Baatz et al. (2015). However, since the method considers only linear scaling of the neutron counts, its application should not alter the calibration accuracy in case of temporally stable above ground biomass (as in this study). Therefore the application makes only sense where temporal biomass dynamics are expected and temporal information on biomass changes are available or in case of cosmic rover applications.

This is true. Once we had analyzed this scenario, the expected outcome (i.e. that it makes no difference) became obvious. We have removed this approach from the entire manuscript. We also tried a different approach by Baroni & Oswald which accounts for seasonally changing biomass in the weighting function. This approach also did not improve our results (probably because the amount of seasonally variable biomass in between the sensor and the soil is not large enough at our filed site).

Specific comments

L143-145: On the web-site of the sensor manufacturer no specification of the measurement technique is given. I suspect that these sensors are actually based on an oscillator-ring as described in Qu et al. (2014) and not on the time-domain transmission technique. In addition, a problem of these sensors could be the top shielding, influencing the soil water content below. Since these kind of sensors only measure soil moisture at a very small volume (only very few centimeters around the sensor blade) this might lead to systematic underestimations of soil moisture.

According to our sensor manual the sensors are indeed based on time-domain transmission. It is hard to find other online information: http://www.ibot.cas.cz/en/kalibrace_stanice_projekt. When calibrating the sensors in the field, we actually found a systematic overestimation at low soil water contents that we corrected for. We did not detect a shielding effect which would cause an underestimation of soil water content.

L147-148: Is the sensor blade actually 15 cm long (at the web site there is no information on the seize and the pictures suggest that the sensor blade to be much shorter)
The sensor blade is actually 16 cm long.

L149: Why didn't you use all data for the calibration?
The sensors were installed in May 2014 only, so we could not use the data for the first 5 calibration campaigns.

L205-208: No scaling needed since this correction considers the relative changes in incoming neutron flux. However, the cutoff rigidity of the Jungfraujoch Station is somewhat different from the study site given is lower latitude. An good choice for the neutron monitor is be the Lomnicky station, Slovakia (LMKS).
We agree that the scaling is unnecessary. Therefore we shortened the paragraph and moved the figure to the supplements. We still use the scaled and gap-filled time series we computed since omitting the scaling does not affect the results.

L206: The correct unit for incoming neutrons is "counts/sec"
We changed all occurrences of 'n h^{-1}' to 'counts h^{-1}'. To stay consistent and avoid confusion we also use hours for the incoming flux corrections.

L224: The methods to determine soil organic carbon and root biomass water equivalents are not presented.
A description of the methods has been added. (See lines 190-195)

L230: This statement is too vague. I think what you meant here is that the objective performance measure is minimized, right?
True. The sentence was removed since we decided to omit the last approach.

L315: What about the other tree species?
We added to the text (Line: 342): 'We did not conduct surveys on the other tree species. Table 1 shows that the beech stand covers 56% of the footprint area around the CRS (when assuming the exponential distance-weighting from Zreda et al. (2008)). Pine covers 16%, spruce 13%, oak 8%. With the new distance weighting function of Köhli et al. (2015), the cover fractions of the other tree species will decrease even further. Also, the seasonal variation in spruce and pine above-ground biomass is very small and thus we consider it to be constant in this study.'

L317: On which grounds did you assume these values?
These values are taken from the literature (papers by Gravano et al. (1999) and Bouriaud et al. (2004)). See line 320.

L359: The correct unit for incoming neutrons is "counts/sec"
Corrected in the revised manuscript.

L380: Change to "the same value for the N0 calibration parameter"
Changed to: 'Following the standard N0-calibration approach of Desilets et al. (2010), we should have ended up with the same N0 value for each of the 10 calibrations.'

L381: The correct unit for the CRS measured neutron intensity is “counts/h”
Corrected in the revised manuscript.

L382-387: According to Zreda et al. (2012) the presence of other hydrogen pools than soil moisture increases the stopping power of the soil, which leads to a change in the slope of the calibration function. Thus, calibration has to be performed using the total hydrogen pool, and soil moisture is then computed by subtracting other hydrogen pools than soil moisture from the measured neutron-derived soil moisture. It is unclear whether this procedure was applied in this study. If not, this would partly explain the differences in soil moisture estimates.
As outlined in the description of our four approaches (esp. Lines 267-285), we used the total hydrogen pool for calibration before subtracting other than soil moisture contributions according to Eq. 4 (Line 240).

L388: The term “new calibration function” is misleading. Changing the “a” parameters of the N0 calibration function is not new and was already presented by Iwema et al. (2015). They called this more adequately “modified N0 method”. However, they only calibrated 3 parameters (the N0 parameter was omitted), because of strong correlations between the parameters leading to ambiguous calibration results (equifinality problem). Did you check for this calibration issue?
You are right that ‘new calibration function’ is misleading. We changed all occurrences to ‘modified calibration function’. It is true that the N0 parameter has a very similar influence on the shape of the calibration function as the a0 parameter. Still we don’t think that in this case equifinality (we agree that it exists) is a real problem for this application because the goal of the calibration simply is to find an efficient function that represents the calibration points. There is no reason to consider any potential adverse implications of the adjusted parameter values or combinations of them. As a side note: we also tested whether simple exponential or gamma functions would perform as well as the 4 parameter calibration function and we found that they in fact did not. So there seems to be justification for the specific set of shapes that is described by the N0-calibration function.

L400-402: Do you have any idea why?
Yes. If you look at Fig. 5 you see that at higher water content a smaller change in neutron counts is associated with a larger change in soil water content (the function is steeper). Therefore the uncertainty during the calibration is also larger.

L414-416: Please provide a figure showing the comparison.
We have prepared a figure and added it as a supplement (Fig S2).

L422-424: Shouldn’t the relationships vary with soil moisture content due changing sensor penetration depths?
Yes, that is what we expected and that is also what we found for the first two approaches. When it is wetter, the penetration depth is reduced for the CRS measurements and the wetter shallower layers receive more weight. Therefore, the CRS measurements show higher SWC than the gravimetrically determined SWC. However, it seems that the distance weighting counters this effect. A probable explanation is that the

formula used for the distance-depth weighting increases the critical depth. This causes higher weights for deeper (drier) soil layers even under wet conditions and could counteract the trend.

L429-430: This finding is quite obvious given the insignificant changes in above biomass. Generally, the application of the vegetation correction makes only sense, when temporal biomass dynamics are expected and temporal information on biomass changes are available.
We removed the whole part on vegetation correction.

L441: This investigation is very similar to Iwema et al. (2015). Please discuss your results in the light of this study.
We added a discussion in section 5.1 (lines 687-695).

L450-457: The results plotted in Fig. 12 show clearly, that only the most extreme dry and wet samplings result in an acceptable calibration result, whereas sampling at intermediate soil moisture will lead to very uncertain calibration of the modified N0-method. On the other hand, this illustrates the value of the standard N0-method that will also produce stable results in case only one sampling date is available. Please add this to the discussion.
Fig. 9 shows that the best 2-point-calibrations are achieved with one sampling point taken under very dry conditions and another sampling point taken either under intermediate or wet conditions. In our case it is hard to see the value of the standard N0-method since it always resulted in too much soil moisture variability no matter whether the calibration was performed during wet, intermediate or dry conditions (because the standard calibration of N0 does not allow a change of the slope of the calibration function).

L458: This chapter belongs to discussion.
We moved parts of this chapter to discussion and only left the parts that really describe results. We also added results on other hydrogen pools, so we renamed the chapter.

L471-474: This is only true when assuming that the CRS footprint is completely covered by beech, which is however not the case.
This is true. So for this calculation, we assume an extreme case since the other vegetation types experience smaller seasonal changes in above-ground biomass. In reality we should expect even less variation in neutron counts due to foliation/defoliation. We will add this statement to the revised manuscript.

L484-485: So the whole discussion of this chapter is unimportant and should be reduced to 1-2 sentences.
Would you say that just because our results suggest that seasonally-varying above-ground biomass does not influence the neutron count significantly the discussion of this finding is not important? We think this finding is very important for the use of CRS in forested areas and worth the extended calculation and discussion. (In the end, it makes life much easier when applying CRS in forests).

L488-518: This section is a summary, not a discussion and thus should be omitted.

We restructured the discussion section.

L520-528: Please discuss your results in the light of the results found by Iwema et al. (2015).
We added a discussion of the findings of Iwema et al. (2015).

L558-560: This statement is not clear to me. Please explain in more detail.
We changed and added more detail: ‘Following the argumentation of Lv et al. (2014), the fact that distance weighting improved our results can be regarded as an indication that non-homogeneous soil moisture conditions indeed lead to changes in the shape of the calibration function. At our site distance weighting reduced the spatial variability within the footprint of the sensor since it assigned higher weights to the closest sampling sites which were all located in the homogenous and relatively wet beech forest, while the influence of the drier soils under the coniferous trees was reduced.’

L564-569: This part is somewhat misleading. Corrections of the neutron count rate (Eqs. 1-3) are essential for any application of the CRS (e.g. Zreda et al., 2012). Vegetation correction is only needed for sites with significant biomass changes. On the other hand, the characterization of the temporal stable hydrogen pools is important for the application of the N0-method. However, the abundance of different pools and the uncertainties in the sensing depth estimation will always lead to uncertainties in the calibration process. As shown by Iwema et al. (2015) and by the results found in this study, this issue can be partly circumvented by the using site specific calibration parameters estimated at using in-situ samples taken during dry and wet conditions. Please reformulate in this sense.
Changed to: ‘If it was possible to fully correct for all factors that influence footprint size, depth-weighting and neutron count, a one-time calibration of the CRS would be sufficient. However, the abundance of different hydrogen pools and the uncertainties in the sensing depth estimation will always lead to uncertainties in the calibration process. Therefore we argue that for using the CRS as a simple tool to measure soil water content at intermediate scales, the efforts of measuring all necessary parameters are not justified. As shown by Iwema et al. (2015) and by the results of this study, this issue can be partly circumvented by using site-specific calibration parameters estimated from in-situ samples taken during dry and wet conditions. Hence, we recommend a two-time calibration that – although being empirical in nature – inherently incorporates many of the required corrections.’.

L583-584: Actually, the seasonal changes of the hydrogen pools in this forest site are negligible. Thus vegetation correction can be omitted.
Agreed.

L594-600: This statement is based on Köhli et al. (2015), but not on results of this study and thus should be omitted.
Since this a very important statement and should be considered by everybody using a CRS, we would like to keep it. But we added the proper reference (Köhli et al.).

L606-608: This step is obvious and should be omitted.

When we first set up the sensor and calibrated it for the first time we brought a battery with us to let the sensor run. After we had finished the soil collection, we took the battery back home with us. Only later it became clear that we should have collected neutron counts for a longer period of time. So we could not use the data from our first calibration effort. This was a hard lesson to learn and we want to make sure that other people do not make the same, admittedly stupid, mistake.

L614: The sampling locations should be adapted to the footprint estimates after Köhli et al. (2015).
We adapted the sampling distances to the footprint estimates after Köhli et al. (2015).

Figures
Figure 1: This map should be integrated in figure 2.
Done.

Figure 1: According to recent results of Köhli et al. (2015) the footprint is considerably smaller than 300 m. Please adapt the figure. In addition, it would be helpful to color the aerial photograph according to the different tree species.
The figure was changed accordingly. Since this figure is also supposed to illustrate our sampling scheme we would like to keep the 'old' footprint size. To make the difference clear we rephrased: 'The yellow circle approximates the footprint of the CRS as it was assumed when sampling took place'. The distribution of different tree species can be seen on Fig. 3 and it would probably make this Figure too busy adding colors or patterns on top of the photograph.

Figure S1: This figure can be omitted (see comment L205-208)
We have moved the Figure to the Supplement (S1).

Figure 2: This schematic figure is wrong in presenting the cosmic-ray neutron intensities as actual rays that are reflected by the soil. The actual processes leading to neutron intensity are far more complex (see e.g. Köhli et al., 2015) and should not be presented in this way in a scientific paper. Also the above ground and below ground footprints are not connected in the simple way as suggested by the schematic drawing. Thus, the figure should be omitted.
We acknowledge the fact that this figure simplifies the actual processes a lot and have added this statement to the caption. We have modified the figure to resolve two of your concerns. The neutron intensities are not any longer depicted as rays and the above-ground footprint was removed entirely. We still think that the figure helps to get an overview over the many parameters that have to be accounted for before/during the use of the CRS method.

Figure 9: The Pareto front needs to be discussed in the text as well.
We realized that it is actually not a Pareto front. So we added to the text: 'The existence of a rather clear front in Fig. 9 indicates that the calibrated neutron count-soil water content conversion will always perform well if the soil moisture differences between the two calibrations are sufficiently large.'

Response to Interactive comment by Anonymous Referee #2

Comments from referee are printed in black. Authors' responses are printed in red.

The authors present a straight forward paper looking at mutli-calibration estimates and methods for a study site in Germany. While the methods in the paper have been previously tested in a variety of ecosystems around the globe the application to this particularly ecosystem is insightful and help further advance the CRNP method. In particular, this site contains a relatively large amount of the total hydrogen in the forest canopy given the sandy and dry soils. The paper is well written and appropriate for the HESS community. My comments and assessment largely follow Heye Bogena so I will only add new comments here or reiterate key points.
We thank reviewer #2 for assessing our manuscript and for providing useful comments on how to improve it.

Major:
The site is interesting in that a potentially a large part of the hydrogen is contained in the biomass instead of the pore water content. A figure like Figure 3 in McJannet 2014 WRR for each calibration data point would be interesting to look at. In addition, perhaps some of the bias in the N0 parameter is because of how hydrogen is more distributed in the clumped biomass instead of distributed throughout the soil. Plotting the differences of N0 vs. relative biomass hydrogen to total hydrogen ratio might show this influence. Franz 2013 GRL supplemental figure S2 illustrated the influence of neutron intensity/counts due to clumped hydrogen in the tree canopy vs. more distributed hydrogen in the soil. Perhaps these detailed calibration datasets might help validate or refute these modeling results.
Very good suggestion. We prepared a figure (similar to the one of McJannet) for 3 distinct soil moisture conditions and added it (Fig. 11). In order to do this we refined our calculations of the different hydrogen pools (now also including SOM, lattice water as well as canopy and litter interception) and also added more detail to the discussion. This new analysis supports the idea that interception can have a significant influence on soil water content measurements performed with a CRS.

The conclusion that the deviation of a single calibration point is upwards of 0.12 m3/m3 is technically correct at the wet end. However, this is a bit misleading given that the neutron counts are never this low or soil moisture this high, particularly at the daily average level, because of the sandy soils. I suggest the authors use the min and max observed counts to properly assess the maximum uncertainty of the method. Looking at Figure 6 it looks like the CRNP never reads above 0.27 m3/m3. All in all, this a fairly small change. Also might be more useful to look at percent absolute error instead of just the difference.
We modified this section stressing the fact that 0.12 m^3/m^3 is the largest deviation that we observed. Due to the sandy soils, the absolute range of soil water content is fairly small at our site. Following your suggestion we changed absolute error to percent absolute error both in the figure and in the text.

Comment: For timescales below the daily level, and thus estimates of the peak soil moisture, clearly some cleaver smoothing filters are needed to estimate the "true peak"

and separate out the signal from the noise. This estimation of the true peak will help constrain things like calculating effective infiltration flux and maybe even runoff depths for water balance studies using the CRNP data.
This is true. We hope that such clever smoothing filters will be developed in the future to provide us with the true peaks using additional information (like time series of precipitation).

The method of determining lattice water by weighing the sample at 105, 400 and 1000 C has not been used by the USA COSMOS community (pg9820 L 18-30). Are they any refs suggesting this is a defensible method compared to the more rigorous approach used by Actlabs? I suspect this difference will be small here as you account for the burn off of carbon. However, for certain soil groups (volcanics?) I imagine this might be problematic. Please add any supporting refs or comment on the pragmatic approach taken here vs. the more rigorous laboratory approaches taken in previous COSMOS work.
The method to determine SOM proposed in this paper is officially called the loss on ignition method. We chose this method since it is indeed pragmatic and does not require the acquisition and handling of toxic chemicals (like for example the Walkley-Black method does). References that describe this method are for example Ball (1964), Ben-Dor & Banin (1989), Davies (1974), Howard & Howard (1990), Schulte et al. (1991). In clay-rich soils the loss on ignition method overestimates SOM content since also some of the lattice water evaporates. This leads to an underestimation of soil hydrogen content (since lattice water contains a higher fraction of hydrogen than soil organic matter). One complication for calcareous soils is the thermal breakdown of carbonates at high temperatures. Although this thermal breakdown can be avoided at temperatures below 430°C (Davies, 1974) at temperatures above that the burned off carbonates would contribute to the lattice water account. We still think that the results obtained with this method provide good estimates of SOM and LW for our purposes with minimal error for most soil groups.

Minor:
P9816 L25. Already is awkward transition. Maybe something like “As early as 1966. . .”
Changed accordingly.

P9837 L8. Franz 2013 WRR investigated the impact of horizontal heterogeneity on the signal.
We are discussing the topic of horizontal heterogeneity in lines 724-739. Here, we are just comparing the count rates of different studies (and in Franz et al. (2013) the lowest count rates are also above 1000 cph).

P9839 L21-25. Again, is this method for lattice water supported by refs? If not then should be noted that this is a pragmatic procedure with expected minimal error for most soil groups other than volcanics, . . . etc. (?). Unfortunately I don’t know all the soil groups this might be affected by so hopefully a pedologist can set us straight.
The method of heating the samples to a temperature of 1000°C to determine lattice water was used in many CRS studies (e.g. Zreda et al., 2012; Bogena et al., 2013). The only complication we found occurs in carbonate-rich soils where thermal breakdown of

carbonates will contribute to the lattice water account. We will add two cautions to the recommendations we give in the appendix.

Response to Interactive comment by Anonymous Referee #3

Comments from referee are printed in black. Authors' responses are printed in red.

The manuscript describes an improved calibration of cosmic-ray soil moisture probes. The problem of calibration has surfaced many times before and different scientists have proposed improvements. But even with those improvements, some problems have remained. This paper adds a valuable component to the growing body of knowledge of how cosmic-ray neutrons respond to surface moisture and how best to calibrate the response functions. The work described in the paper is thoughtfully designed and carefully executed and analyzed. The conclusion is simple, but important: calibration function is different than that originally proposed, and that difference can be captured by doing only two calibrations at two different levels of soil moisture. The good quality of work and the important conclusion make the paper important. I recommend that it be published with minor revisions suggested on the annotated copy (attached).
Thanks to reviewer #3 for a considerate and helpful review that helped in improving the paper.

Line 9: this is instrument, not technique. The technique is "measuring soil moisture with cosmic-ray neutrons".
Changed to: 'Measuring soil moisture with cosmic ray neutrons is a promising technique for intermediate spatial scales.'

Line 12: Desilets' function was computed using MCNPX
Changed to: 'The calibration is based on soil water content derived directly from soil samples taken within the footprint of the sensor.'

Line 21: this conclusion is based on ten calibration data sets, but these ten sets span a narrow range of average soil moisture values between 7% and 16% (Fig. 3) - so how solid is this conclusion?
Soil moisture in our very sandy soils only ranges between 5% and 25%. So although 16% still appears to be on the dry end (when comparing it to more silty/clayey soils) we are covering 50% of the total range of possible soil moisture. Moreover, it will be difficult to calibrate at even higher soil water contents since they usually only appear for a short period of time after large precipitation events. At that time, however, conditions for calibration are not optimal since there will still also be intercepted water in the canopy and litter layer. Also, keep in mind that the values on Fig. 3 are non-weighted averages of all 108 soil samples. The weighted averages used for calibration range from 8% to 22% (see Table 4, column: θ_{depthW}). We added some of this information to the manuscript in chapter 4.1.

Line 32: meaning? seems unnecessary
Agreed. Removed.

Line 34: the largest spatial variance is usually at intermediate moisture levels, and decreases towards both dry and wet ends
Changed to: '…especially under intermediate wetness conditions…'

Line 37: delete "very". perhaps rephrase this to provide more specific information, for example: "this is time-consuming and expensive"
Changed to: 'This can be time-consuming and expensive.'

Line 45: not from the sun (not energetic enough to produce cascade neutrons); say something like "Cosmic-ray neutrons on Earth are formed by high-energy protons coming from galactic sources, such as supernovae." cascade neutrons are high-energy neutrons, not fast neutrons. Fast neutrons are produced by secondary neutrons via the process of evaporation. So the fast neutrons are tertiary neutrons, by that term is not used; term evaporation neutrons is used instead.
Thanks for this clarification. This is quite complicated. Changed to: 'Cosmic ray neutrons on Earth are formed when high-energy protons deriving from galactic sources (such as supernovae) enter the Earth's atmosphere. Once in the atmosphere, the protons interact with atomic nuclei (mainly nitrogen and oxygen) producing cascades of secondary neutrons (also called high-energy neutrons) that travel towards the Earth's surface and into the soils. When secondary neutrons interact with air or soil they trigger the release (evaporation) of fast (but low-energy) neutrons.'

Line 61: specify that Kodama's detector was buried in the soil.
Added: '…with a neutron sensor buried in the soil.'

Line 63: larget than what? perhaps say "large" or specify that area (~30 ha).
Added: '…(~30 ha)…'.

Line 68: this is not appropriate reference; the range 10 cm - 70 cm was given in Zreda et al., 2008. Also, the dependence on soil water content was specified, but not dependence on soil type and distance from sensor. Koehli et al., 2015 worked out the dependence on distance.
Exchanged the references.

Line 102: why unwanted? perhaps better to say "adds additional signal" that complicates analysis (or complicates extraction of soil moisture signal).
Changed to: '…and adds additional temporal variability to the CRS time series complicating the extraction of the soil moisture signal.'

Line 108: please make sure that this is true; I think the idea is to fix the calibration parameter at the high end (water), rather than low end (dry soil, in Desilet's calibration function), but that is a regression parameter, not measured value.
In the 2013 paper they say that this parameter can be easily retrieved from measurements over a large water body. We think you are right in noting that it is in fact a regression parameter. So we changed the wording to make this clearer.

Line 135: Are they applicable also at the site? 1.6 km seems close, but precipitation and humidity can vary at smaller scales.
We compared the climate data with data we collected ca. 400 m away from our site and the differences were marginal. We decided to use the data from 1.6 km since it was a longer data set (we only installed the 400 m sensors in the middle of 2014).

Line 175: and assuming the sensitivity decrease with distance given in Zreda et al., 2008, for the footprint radius of ca 300 m. The new footprint estimate by Koehli et al., 2015, gives smaller footprint, so the sampling distances for equal weights will be smaller than your 25 m, 75 m and 200 m.

This is true. We now discuss this in more detail later in the manuscript (see Section 4.2):' Sampling distances with equal weights according to Köhli et al. (2015) would have differed from our sampling pattern (~1 m, ~33 m, ~140 m instead of 25 m, 75 m, 200 m), a condition which we balance by adjusting the distance weights. Furthermore the conditions within 30 m around our CRS are quite homogenous since the sensor is located within a pure beech stand and we are expecting little difference in soil moisture content between locations at 1 and 25 m distance.'

Line 204: this correction uses shielding depths (g/cm2), not pressure (hPa); however, the difference in this case would be cosmetic (or zero, if the value of 133.3 hPa was obtained from the equivalent shielding depth), so it does not matter for the results; on the other hand, if you want to be consistent with cosmic-ray literature, please make the change to shielding depth.

We changed the units to shielding depth. Thanks for providing us with the conversion factor.

Line 207: reference Zreda et al., 2012 for this correction

Added the reference.

Line 216: this difference (309 vs 327) cannot possibly be due to difference in altitude between the monitors (which should be a factor of four or more). I think these data sets are pressure corrected, and the difference idiosyncrasies of the two neutron monitors.

Deleted the mention of altitude.

Line 219: reference Rosolem et al., 2013 for this correction

Reference to Rosolem et al., 2013 is now given.

Line 320: is this 0.6 kg of H_2O per kg of dry biomass? or wet biomass? clarify.

It is actually 0.6 kg of H_2O per kg of wet biomass. We clarified this in the manuscript.

Line 321: this seems low. does this figure include free water (H_2O in xylem) and cellulose-bound water (OH group)?

No, this just includes free water. The cellulose-bound water is calculated in the next step (assuming hydrogen content in dry biomass is 0.0622 kg kg-1).

Line 335: isn't there a seasonality or some other temporal variability (eg, with droughts) in the free water within trunks and branches?

Yes, there is some seasonality also in this regard (there is even daily fluctuations indicating variations in transpiration flux). This, however, was not part of our analysis since it is difficult to determine the exact numbers and it is likely that these variations are too small to influence the neutron count.

Line 342: This section should provide more information, not just the equation. What is KGE? How is it used? What is compared with what? How are these variables computed (eg, r)? What is the significance of the result? Etc...

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

We added the equation for the correlation coefficient r and some more information on the KGE': 'The KGE' measures the Euclidian distance in a 3-D space where the correlation coefficient r is on one axis, the variability ratio β is on the second axis and the bias ratio γ is on the third axis. KGE' scores range from 1 (representing a perfect fit) to -∞. Due to the composite nature of the KGE' it is relatively simple to analyze which feature of the time series (correlation, bias, variability) contributes most to the good/bad performance of a model.'

Line 398: this is inaccurate; one time series shown in Fig. 7 has differences of that magnitude, whereas the other nine have considerably smaller deviations; please, state this result correctly, without creating undue alarm.
You are right in that our formulation is unclear and could be interpreted in different ways. So we will modify the statement: 'As a consequence, the 10 computed time series based on the standard N0-calibration function of Desilets et al. (2010) also showed differences in volumetric soil water content (Fig. 4 illustrates results for the DDW approach). In the most extreme case, these differences were larger than 0.1 m3 m-3 (which is equal to 30 % of the total range of soil water content at the site).'.

Line 403: It is unclear how this conclusion was reached. Please describe in detail what was done, what were the results and how to interpret them.
We conclude that there is no unique calibration parameter and describe the calibration procedure in detail in the methods section (lines: 256-303). The results are presented in Table 5 and Fig. 4 & 5. We will add a more detailed interpretation of the results to the discussion section.

Line 420: There is a lot of noise in Figure 5 and the difference between standard N0 and the improved data is not clear. Can you add error bars to the data points?
We are not sure about this comment. In our view, this difference between the standard and the modified calibration functions is clearly represented by the dotted and the solid lines respectively. Since there is already a lot going on in this figure, we decided not to add error bars and we were not sure of the added value they would provide.

Line 553: The absolute count rate has no influence on the shape of the response function, just on the precision of calibration. I would remove this conclusion or reword it to make this conclusion ("it clearly does not at our site") less strong.
Ok. We removed this conclusion.

Line 564: but on page 9831 (of my copy) you stated that the sensitivity is better when using your calibration. Please, make these two statements consistent.
On page 9831 we stated that the sensitivity of the sensor is essentially higher than it should be (not better). This means that already a small difference in neutron counts indicates a large difference in soil moisture. The modified calibration accounts for this by

decreasing the slope of the calibration function and thereby reducing the sensitivity of the sensor (so that now a bigger difference in neutron counts is required to cause differences in the soil water content reading). We modified our statement on page 9831 (= manuscript line 475) to: '…essentially the sensor has a higher resolution/sensitivity than one would expect.'

Line 642: Franz et al 2012 is not the correct reference; Franz et al merely repackaged the information given in Zreda et al 2012.

We decided to recommend sampling distances according to Köhli et al. (2015). Concerning the sampling pattern we used: Franz et al. 2012 describe 3 circles with distances around the CRS of 25, 75 and 200 m. Zreda et al. 2012 describe 3 circles with distances around the CRS of 25, 75 and 175 m. Since we used 25, 75 and 200 m the more correct citation to describe our calibration setup would be Franz et al. 2012. However, to recognize the contribution of Zreda et al. we inserted the reference in line 174: '…we followed the recommended sampling pattern for the calibration of CRS which was developed by Zreda et al. (2012) and slightly modified and detailed in Franz et al. (2012b).'.

Table 4: Insert line!

I would love to. Actually there was a line when we submitted the manuscript. But it seems that HESS does not allow lines within tables. However, I will try to convince them to leave this line in there because I also think it is necessary.

Table 5: I am not sure how to read these results. Can you clarify in the figure caption or in text (which also glosses over this in one short paragraph at the end of section 4.3).

We modified the text to make the table clearer: 'All resulted in largely deviating N0-values between the individual calibrations (see means and standard deviations in column 1 and 2 of Table 5). This in turn led to differences in the time series of volumetric soil water content between the individual calibrations (see means and standard deviations in column 3 and 4 of Table 5).'

Fig. 2: can you group these factors into two groups: (1) constant in time (eg, WL), and (2) variable in time (eg, water vapor)? And then perhaps also into another groups: (3) those measured at calibration (eg, WL), and (4) those measured as time series (eg, water vapor)? The significance is that some parameters are easy to handle, others are difficult, and the two should not be commingled.

We revised the figure according to your suggestions. We grouped the factors into two groups and mentioned which of them need to be monitored continuously and which can be accounted for during calibration.

Use of cosmic ray neutron sensors for soil moisture monitoring in forests

Ingo Heidbüchel[1], **Andreas Güntner**[1], **Theresa Blume**[1]

[1] {GFZ German Research Centre for Geosciences, Helmholtz Centre, Potsdam, Germany}

Correspondence to: I. Heidbüchel (ingo.heidbuechel@gfz-potsdam.de)

Abstract

Measuring soil moisture with c~~C~~osmic ray neutrons ~~sensors (CRS)~~ is~~are~~ a promising technique ~~to measure soil moisture at~~ for intermediate spatial scales. To convert neutron counts to average volumetric soil water content a simple calibration function can be used (the N_0-calibration of Desilets et al., 2010). The~~is~~ calibration ~~function~~ is based on soil water content derived directly from soil samples taken within the footprint of the sensor. We installed a cosmic-ray neutron sensor (CRS) in a mixed forest in the lowlands of north-eastern Germany and calibrated it 10 times throughout one calendar year. Each calibration with the N_0-calibration function resulted in a different CRS soil moisture time series, with deviations of up to 24 % of the total range~~0.12 m^3 m^{-3}~~ for individual values of soil water content. Also, many of the calibration efforts resulted in time series that could not be matched with independent in situ measurements of soil water content. We therefore suggest a modified~~new~~ calibration function with a different shape that can vary from one location to another. A two-point calibration proved to be adequate to correctly define the shape of the modified~~new~~ calibration function if the calibration points were taken during both dry and wet conditions ~~covering~~ spanning at least ~~50 %~~half of the total range of soil moisture. The best results were obtained when the soil samples used for calibration were linearly weighted as a function of depth in the soil profile and non-linearly weighted as a function of distance from the CRS, and when the depth-specific amount of soil organic matter and lattice water content was explicitly considered. The annual cycle of tree foliation was found to be a

Kommentar [A1]: Rev3: this is instrument, not technique. The technique is "measuring soil moisture with cosmic-ray neutrons"

Kommentar [A2]: Rev3: Desilets' function was computed using MCNPX

Kommentar [A3]: Rev3: this is very important for users

Kommentar [A4]: Rev3: this conclusion is based on ten calibration data sets, but these ten sets span a narrow range of average soil moisture values between 7% and 16% (Fig. 3) - so how solid is this conclusion?

Kommentar [A5]: Reply: Soil moisture in our very sandy soils only ranges between 5% and 25%. So although 16% still appears to be on the dry end (when comparing it to more silty/clayey soils) we are covering 50% of the total range of possible soil moisture. Moreover, it will be difficult to calibrate at even higher soil water contents since they usually only appear for a short period of time after large precipitation events. At that time, however, conditions for calibration are not optimal since there will still also be intercepted water in the canopy and litter layer. Also, keep in mind that the values on Fig. 3 are non-weighted averages of all 108 soil samples. The weighted averages used for calibration range from 8% to 22% (see Table 4, column: θ_{depthW}). We added some of this information to the manuscript in chapter 4.1.

negligible factor for calibration because the variable hydrogen mass in the leaves was small compared to the hydrogen mass changes by soil moisture variations. As a final point~~Finally~~, we provide a best practice calibration guide for CRS in forested environments.

1. Introduction

~~Measuring~~ Determining average soil moisture content ~~comprehensively~~ over larger areas is difficult, mainly for two reasons. Firstly, soil moisture can be highly variable ~~already~~ even at small spatial scales, especially under intermediate~~dry~~ wetness conditions (e.g. Western et al., 2004). Secondly, most common in situ measurement techniques only yield point measurements. To obtain a valid estimate of area-average soil moisture one needs to collect data from numerous locations within a given area. This can be ~~very~~ time-consuming and expensive~~cumbersome~~. More recently, remote sensing of soil moisture at larger scales has become a research focus (e.g. see Ochsner et al., 2013 for a recent review); however, ~~up to this point~~ the measurement depth of many of these methods is still limited to the upper 5 cm of the soil. Also, both spatial and temporal resolution is rather coarse. A technique that intends to bridge the scale gap between point measurements of soil moisture and remote sensing is the use of cosmic ray neutrons as indicators of soil moisture. A detailed description ~~of the functioning~~ of the cosmic ray neutron sensors (CRS) can be found in Zreda et al. (2008, 2012), here we will only describe the basic measurement principle. Cosmic ray neutrons on Earth are formed when high-energy protons~~cosmic ray particles~~ deriving from ~~the sun (and also from other~~ galactic~~xies)~~ sources (such as supernovae) enter the Earth's atmosphere. Once in the atmosphere~~there~~, the protons~~y~~ ~~start interacting~~interact with ~~other~~ atomic nuclei (mainly nitrogen and oxygen) producing cascades of secondary~~fast~~ neutrons (~~that are~~ also called high~~low~~-energy neutrons) that travel towards the Earth's surface and into the soils. When secondary neutrons interact with air or soil they trigger the release (evaporation) of fast (but low-energy) neutrons. The number of fast neutrons above the soil surface depends strongly on the number of hydrogen atoms in the surroundings because hydrogen atoms have a very high capacity to moderate fast cosmic ray neutrons (that means to slow them down and turn them into thermal neutrons with even less energy – effectively removing the fast neutrons from the system). The number of hydrogen atoms increases with increasing soil water content and hence soils with high water contents re-emit

Kommentar [A6]: Rev3: meaning? seems unnecessary

Kommentar [A7]: Rev3: the largest spatial variance is usually at intermediate moisture levels, and decreases towards both dry and wet ends

Kommentar [A8]: Rev3: delete "very"

perhaps rephrase this to provide more specific information, for example: "this is time-consuming and expensive"

Kommentar [A9]: Rev3: not from the sun (not energetic enough to produce cascade neutrons); say something like "Cosmic-ray neutrons on Earth are formed by high-energy protons coming from galactic sources, such as supernovae."

Kommentar [A10]: Rev3: cascade neutrons are high-energy neutrons, not fast neutrons. Fast neutrons are produced by secondary neutrons via the process of evaporation. So the fast neutrons are tertiary neutrons, by that term is not used; term evaporation neutrons is used instead.

fewer fast neutrons than soils with low water content. That leads to fewer fast neutrons being detected above-ground by the CRS which is generally installed 1-2 m above the soil surface.

~~Already in~~As early as 1966 Hendrick and Edge reported that the intensity of fast (low-energy) neutrons (~1 keV) detected above the~~-~~ ground depended on the hydrogen content of the soil, and Kodama (1985) found an inverse correlation of neutron intensity and soil moisture content w~~hen~~ith a ~~burying a~~ neutron sensor buried in the soil. In 2008, Zreda et al. introduced a method to measure average soil water content over a larger area (~30 ha) with CRS. The footprint of CRS, i.e. the area around the sensor where 86 % of detected neutrons originate from, covers a circle with an approximate radius of 300 m (Desilets and Zreda, 2013). However, the radius can decrease with increasing air density and humidity, with increasing vegetation density and with increasing soil moisture to about 100 m (Köhli et al., 2015). The ~~critical~~effective measurement depth of CRS, i.e. the soil depth where 86 % of detected neutrons originate from, varies between 10 and 70 cm below surface (Zreda et al., 2008~~Franz et al., 2012a~~), depending on soil type, water content and distance from the sensor (Köhli et al., 2015). To account for the contributions of neutrons from different soil depths, various~~Different~~ depth-weighting approaches have been proposed, some of them assuming a linear decrease of weights with depth (Franz et al., 2012a), others assuming a non-linear decrease with depth (Köhli et al., 2015).

Kommentar [A11]: Rev2: Already is awkward transition. Maybe something like “As early as 1966.”

Kommentar [A12]: Rev3: specify that Kodama's detector was buried in the soil.

Kommentar [A13]: Rev3: larger than what? perhaps say "large" or specify that area (~30 ha).

Kommentar [A14]: Rev3: this is not appropriate reference; the range 10 cm - 70 cm was given in Zreda et al., 2008. Also, the dependence on soil water content was specified, but not dependence on soil type and distance from sensor. Koehli et al., 2015 worked out the dependence on distance.

The original measurement method uses a relationship between neutron flux and volumetric soil water content with the shape of the relationship being known from neutron transport simulations. For this relationship, Desilets et al. (2010) presented an equation with three constant shape parameters (a_0, a_1, a_2) and one calibration parameter (N_0) which has to be calibrated with soil moisture values determined by the gravimetric method from field soil samples. The influence of soil lattice water and soil organic matter on the signal was investigated by Zreda et al. (2012). They found that both lattice water and soil organic matter contain fixed amounts of hydrogen that further attenuate the neutron signal and need to be taken into account. Lattice water and soil organic matter corrections to the original relationship by Desilets et al. (2010) are provided for example~~e.g.~~ in Lv et al. (2014).

Other external factors influencing the neutron count that need to be corrected for are (a) atmospheric pressure (Bachelet et al., 1965), (b) incoming neutron flux (see e.g. Zreda et al., 2012, Bogena et al., 2013) and (c) specific humidity (Rosolem et al., 2013). More recently, the effects of ~~living~~ biomass on the neutron signal have been discussed. Bogena et al. (2013) noted that aboveground biomass reduced the neutron count rate and thus decreased the sensitivity of the sensor. To counter this loss of sensitivity they recommended a 24 h integration time for their forested catchment as a compromise between decreased uncertainty and decreased time resolution. Hawdon et al. (2013) and Baatz et al. (2015) compared neutron counts for locations with different amounts of biomass. Hawdon et al. (2013)~~and~~ reported that the variation in biomass could explain 80 % of the variation in neutron counts when assuming a nonlinear relationship between biomass and neutron counts,~~.~~ Baatz et al. (2015) ~~also related biomass to neutron counts, but~~ explained 87 % of the variation proposing~~ed~~ a linear relationship between the two variables. Baroni and Oswald (2015) suggested that the influence of above-ground biomass between the sensor and the ground which decreases the ~~critical~~effective measurement depth of the CRS can be incorporated into the weighting approach of Franz et al. (2012a). This is especially important in locations where frequent large biomass changes occur, for example in agricultural fields. Coopersmith et al. (2014) found that soil moisture in a corn crop is often overestimated when the leaf area index (LAI) is relatively high while it is underestimated when LAI is relatively low – circumstances which could cause differences in the calibration and resulting soil moisture measurements ~~between the seasons~~. The influence of the litter layer in forested environments was investigated by Bogena et al. (2013). Water content in the litter layer changes rapidly and adds additional~~unwanted~~ temporal variability to the CRS time series complicating the extraction of the soil moisture signal. Therefore, Bogena et al. (2013) recommended considering the water dynamics in the litter layer explicitly in the calibration approach. Franz et al. (2013) introduced a new approach (the universal calibration function) that takes into account all sources of hydrogen thereby requiring estimates of lattice water, soil organic carbon, and vegetation biomass as well as a ~~calibration~~ regression factor that can be derived from calibration or ~~also be~~may directly be retrieved from neutron count measurements over a large water body (500 m on all sides and deeper than 1 m).

Since the launch of the cosmic ray neutron method many changes and corrections have been brought forward that altered the way the method is applied. These changes and corrections can be divided into two groups. On the one hand, there are corrections that are applied to the raw neutron count in order to remove the influence ~~that~~ of other variables ~~have on the signal~~ (such as air pressure and humidity variations or fluctuations in incoming neutron counts). On the other

Kommentar [A15]: Rev3: why unwanted? perhaps better to say "adds additional signal" that complicates analysis (or complicates extraction of soil moisture signal).

Kommentar [A16]: Rev3: please make sure that this is true; I think the idea is to fix the calibration parameter at the high end (water), rather than low end (dry soil, in Desilet's calibration function), but that is a regression parameter, not measured value.

Kommentar [A17]: Reply: In the 2013 paper they say that this parameter can be easily retrieved from measurements over a large water body. We think you are right in noting that it is in fact a regression parameter. So we changed the wording to make this clearer.

hand, changes have been made to the way we average the soil moisture measurements during the calibration campaigns in order to get a representative soil moisture value that corresponds to what the sensor actually “sees” at the time of calibration (changing ~~critical~~effective measurement depth, changing footprint diameter, inclusion of lattice water and soil organic matter water equivalent). All this has led to improvements in the method’s accuracy for many environments. Most of these studies were performed in medium to high-count environments with neutron count rates above 1000 counts per hour, in generally dry environments, at higher elevations and with little vegetation. Only a few studies were performed in low-count environments with count rates below 1000 counts per hour (e.g. Rivera Villareyes et al., 2011; Bogena et al., 2013). In the present study, we evaluated whether the CRS also provides reliable and consistent soil moisture measurements in a low-count environment, i.e., in a temperate mixed forest close to sea level. We tested several weighting approaches to convert gravimetrically determined soil water content of the top 30 cm into an average soil water content that can be used for the calibration of the CRS. Additionally, we analyzed whether the annual forest cycle of foliation and defoliation is important to consider for instrument calibration. ~~Finally, we~~ We furthermore compiled a best-practice for the calibration of CRS in forested, low-count environments which is provided in Appendix A.

2. Field site and instrumentation

The CRS (CRS-1000 by Hydroinnova) was installed in late 2013 in the Müritz National Park in north-eastern Germany (53°19'49.0"N, 13°11'56.5"E) at an elevation of about 84 m a.m.s.l. (Fig. 1, inset). Precipitation, temperature and relative humidity data was provided by the climate station Serrahn (1.6 km to the north). Average annual air temperature at the site is 8°C with a maximum in July (17.2°C) and a minimum in January (-0.9°C). Average annual precipitation is 580 mm with a maximum in June (65 mm) and a minimum in February (28 mm). This makes for a maritime temperate climate (Cfb) in the Köppen climate classification. The sensor is located in a sandy outwash plain, a relic from the last glaciation, which causes the soil texture to be ~~relatively~~ homogeneous with sand fractions of about 95% throughout the entire profile. Data from a nearby well shows that the groundwater level at the site is almost 20 m below the terrain surface. The vegetation within the sensor footprint consists of both deciduous and coniferous

Kommentar [A18]: Rev3: Are they applicable also at the site? 1.6 km seems close, but precipitation and humidity can vary at smaller scales.

Kommentar [A19]: Reply: We compared the climate data with data we collected ca. 400 m away from our site and the differences were marginal. We decided to use the data from 1.6 km since it was a longer data set (we only installed the 400 m sensors in the middle of 2014).

trees. Immediately surrounding the sensor is a mature beech forest (*Fagus sylvatica* L., older than 100 years), also within the footprint (but farther away) with a distance of at least 40 m from the sensor there is young pine (*Pinus sylvestris* L.), oak (*Quercus robur* L.) and spruce (*Picea abies* (L.) H.Karst.) forest (all younger than 50 years) as well as a small strip of open grassland (see Fig. ~~2~~1 and also ~~Fig. 6~~Fig. 3 for a map of the forest stands and Table 1 for fractions of the different tree stands within the footprint). Depending on the tree species, the mineral soil is covered by an organic soil layer and a litter layer of variable depth and water holding capacity.

For validation of the CRS soil water content measurements, in May of 2014 we installed 18 soil moisture sensors (TOMST) close to the soil sampling/calibration locations. They are based on the principle of time domain transmission (TDT) and each sensor comes with its own logger and power supply (more information under: http://www.tomst.cz/tms/TMS-3.html). These sensors were installed vertically from the terrain surface into the soil so that they continuously measure soil water content averaged over the top ~~15~~ 16 cm of the soil. In order to calibrate the sensors we used the gravimetric soil moisture data we collected from the upper 15 cm during the last five calibration campaigns which were carried out within the measurement period of the sensors (June-November 2014)~~(SU, F1-F4)~~. The volumetric water content within the upper 15 cm of the CRS footprint was calculated as the mean of all 18 TDT sensors.

3. Methods

3.1. Calibration

We conducted a total of 10 calibration campaigns throughout one calendar year (2014). The first one (WI) took place in February during winterly conditions with very wet soils. The next four calibrations (S1-4) followed in spring (April-May) and covered the entire period of tree foliation. The sixth calibration (SU) was done under very dry conditions in July and the last four calibrations (F1-4) in fall (October-November) covering the trees' defoliation. For all the calibration campaigns we followed the recommended sampling pattern for the calibration of CRS which was developed by Zreda et al. (2012) and slightly modified and detailed in Franz et al. (2012b). The sampling pattern prescribes 3 concentric circles around the CRS with radii of 25, 75

Kommentar [A20]: Rev1: On the web-site of the sensor manufacturer no specification of the measurement technique is given. I suspect that these sensors are actually based on an oscillator-ring as described in Qu et al. (2014) and not on the time-domain transmission technique. In addition, a problem of these sensors could be the top shielding, influencing the soil water content below. Since these kind of sensors only measure soil moisture at a very small volume (only very few centimeters around the sensor blade) this might lead to systematic underestimations of soil moisture.

Kommentar [A21]: Reply: According to our sensor manual the sensors are indeed based on time-domain transmission. It is hard to find other online information: http://www.ibot.cas.cz/en/kalibrace_stanice_projekt. When calibrating the sensors in the field, we actually found a systematic overestimation at low soil water contents that we corrected for. We did not detect a shielding effect which would cause an underestimation of soil water content.

Kommentar [A22]: Rev 1: Why didn't you use all data for the calibration?

Kommentar [A23]: Reply: The sensors were installed in May 2014 only, so we could not use the data for the first 5 calibration campaigns.

Kommentar [A24]: Rev1: Is the sensor blade actually 15 cm long (at the web site there is no information on the seize and the pictures suggest that the sensor blade to be much shorter)

Kommentar [A25]: Reply: The sensor blade is actually 16 cm long.

and 200 m, respectively (Fig. ~~2~~1). The 3 circles are intersected by 6 straight lines that point from the sensor towards north (0°), north-east (60°), south-east (120°), south (180°), south-west (240°) and north-west (300°). Samples are taken in the vicinity of all intersections – the samples do not have to be taken at the exact spot of the intersection. This sampling pattern ensures that each sample has equal weight towards the spatial mean of soil moisture that is detected by the CRS, assuming that the sensitivity of the CRS decreases exponentially with distance. We used a split-tube sampler to extract 30 cm soil cores at 18 locations within the footprint of the sensor afterwards dividing each soil core into six 5 cm thick soil samples. For each of the 10 calibrations this left us with 108 soil samples which were then transferred in sealed plastic bags to the laboratory where they were immediately weighed, then oven-dried at 105°C for 24 h and then weighed again to determine their volumetric water content and bulk density. Afterwards, lattice water, soil organic matter content and root biomass were determined for six depth-representative soil samples. To this end the 108 samples (taken from the last calibration campaign in November) were grouped by sampling depth. We extracted 2 g from each of the~~all~~ 18 samples ~~from one~~per sampling depth and combined them to create one bulk sample per depth. Then, the already oven-dried samples were weighed and put in the oven for another 24 h at a temperature of 400°C. The procedure is called ‘loss on ignition’ since the organic matter is burned off during the process (Ball, 1964; Davies, 1974). This removed most of the soil organic matter and root biomass from the samples. After weighing the samples (to compute the fraction of combined soil organic matter and root biomass) they were again placed in the oven for 24 h, this time at a temperature of about 1000°C. After that, the lattice water was also removed from the samples. A final weighing yielded the fraction of lattice water per soil depth. In order to make soil organic matter and root biomass comparable to the influence of pure water we converted them into equivalents of water by multiplying their weight by 0.556 which is the ratio of five times the molecular weight of water to the molecular weight of cellulose (taking into account that cellulose ($C_6H_{10}O_5$) contains 10 hydrogen atoms per molecule while water (H_2O) only contains two) (Hawdon et al., 2014).

Kommentar [A26]: Rev3: and assuming the sensitivity decrease with distance given in Zreda et al., 2008, for the footprint radius of ca 300 m. The new footprint estimate by Koehli et al., 2015, gives smaller footprint, so the sampling distances for equal weights will be smaller than your 25 m, 75 m and 200 m.

Kommentar [A27]: Reply: This is true. We now discuss this in more detail later in the manuscript (see Section 4.2):’ Sampling distances with equal weights according to Köhli et al. (2015) would have differed from our sampling pattern (~1 m, ~33 m, ~140 m instead of 25 m, 75 m, 200 m), a condition which we balance by adjusting the distance weights. Furthermore the conditions within 30 m around our CRS are quite homogenous since the sensor is located within a pure beech stand and we are expecting little difference in soil moisture content between locations at 1 and 25 m distance.’

Kommentar [A28]: Rev3: 2 g from each of the 18 samples?

The neutron counts from the sensor were smoothed with a 12 h moving window to reduce measurement noise (see Bogena et al., 2013). The next step was to correct the neutron counts for variations in (a) pressure, (b) incoming neutron flux and (c) water vapor in the air. This was done by applying the following corrections:

a. Pressure correction:

$$N_p = N_{raw} * e^{\left(\frac{x\cancel{P} - x\cancel{P}_0}{L}\right)}$$

(1),

with N_p being the pressure corrected neutron counts (~~n h^{-1}~~counts h^{-1}), N_{raw} the raw neutron counts (~~n h^{-1}~~counts h^{-1}), ~~*P*~~*x* the atmospheric shielding depth~~pressure~~ (g cm^{-2}) for every time step (derived from atmospheric pressure measured directly inside the CRS case)~~ for every time step (hPa)~~, ~~*P*~~x_0 the average atmospheric shielding depth~~pressure~~ (g cm^{-2}~~hPa~~) for the entire measurement period and *L* the effective nucleon attenuation length for high-energy neutrons (for our site we assumed a value of 135.9 g cm^{-2} which is equivalent to 133.3 hPa) (Desilets and Zreda, 2003). To convert atmospheric pressure (hPa) into shielding depth (g cm^{-2}) the atmospheric pressure has to be multiplied by 1.0194 s^2 m^{-1}.

b. Incoming flux correction (Zreda et al., 2012):

$$N_{pi} = N_p * \frac{N_{avg}}{N_{nm}} \qquad (2),$$

with N_{pi} being the sensor neutron count rate corrected for changes in atmospheric pressure and incoming neutrons (~~n h^{-1}~~counts h^{-1}), N_{avg} the average count rate of incoming neutrons (~~n h^{-1}~~counts h^{-1}) over the entire measurement period and N_{nm} the neutron count rate of the neutron monitor for each time step (~~n h^{-1}~~counts h^{-1}).

As the time series of the closest neutron monitor, located in Kiel, Germany, contains several data gaps, we selected the continuous time series of the Jungfraujoch, Switzerland, for this study. We scaled this time series by adjusting its mean (309 ~~n/h~~counts h^{-1}) to the mean of the Kiel time series (327 ~~n/h~~counts h^{-1})~~ in order to account for the difference in altitude and latitude between the two neutron monitors~~. The resulting time series resembles the Kiel time series very closely (Fig. S1~~3~~).

c. Water vapor correction (Rosolem et al., 2013):

$$N_{pih} = N_{pi} * \left[1 + 0.0054 * \left(p_{v0} - p_{v0}^{ref}\right)\right] \qquad (3),$$

Kommentar [A29]: Rev3: this correction uses shielding depths (g/cm2), not pressure (hPa); however, the difference in this case would be cosmetic (or zero, if the value of 133.3 hPa was obtained from the equivalent shielding depth), so it does not matter for the results; on the other hand, if you want to be consistent with cosmic-ray literature, please make the change to shielding depth.

Kommentar [A30]: Reply: We changed the units to shielding depth. Thanks for providing us with the conversion factor.

Kommentar [A31]: Rev3: reference Zreda et al., 2012 for this correction

Kommentar [A32]: Rev1: The correct unit for incoming neutrons is “counts/sec”

Kommentar [A33]: Reply: Changed all occurrences of ‘n h^{-1}’ to ‘counts h^{-1}’. To stay consistent and avoid confusion we also used hours for the incoming flux corrections.

Kommentar [A34]: Rev3: this difference (309 vs 327) cannot possibly be due to difference in altitude between the monitors (which should be a factor of four or more). I think these data sets are pressure corrected, and the difference idiosyncrasies of the two neutron monitors.

Kommentar [A35]: Rev1: No scaling needed since this correction considers the relative changes in incoming neutron flux. However, the cutoff rigidity of the Jungfraujoch Station is somewhat different from the study site given is lower latitude. An good choice for the neutron monitor is be the Lomnicky station, Slovakia (LMKS).

Kommentar [A36]: Reply: We agree that the scaling is unnecessary. Therefore we shortened the paragraph and moved the figure to the supplements. We still use the scaled and gap-filled time series we computed since omitting the scaling does not affect the results.

Kommentar [A37]: Rev3: reference Rosolem et al., 2013 for this correction with N_{pih} being the sensor neutron count corrected for changes in pressure, incoming neutrons and water vapor (~~n h^{-1}~~counts h^{-1}), p_{v0}^{ref} the average absolute humidity of the air over the entire measurement period (g m^{-3}) and p_{v0} the absolute humidity for each time step (g m^{-3}). The constant 0.0054 has units of m^{3} g^{-1}.

Finally, to convert corrected neutron counts (N_{pih}) into volumetric soil moisture (θ), Desilets et al. (2010) introduced an equation with four parameters – three of which (a_0 = 0.0808, a_1 = 0.372, a_2 = 0.115) were determined via neutron transport simulations and a fourth one (N_0) that serves as a calibration parameter accounting for site and sensor specific variations and representing neutron counts over dry soil at reference conditions during calibration:

$$\theta(t) = \left\{\left[a_0 * \left(\frac{N_{pih}(t)}{N_0} - a_1\right)^{-1} - a_2\right] * \rho_{bd}\right\} - W_L - (SOM + B_R) \quad (4).$$

The other parameters ρ_{bd}, W_L, SOM and B_R can be measured directly from the calibration soil samples: the bulk density of the soil (ρ_{bd} in g cm^{-3}), the summed volume fraction of lattice water in the soil grains and tightly bound water (W_L in m^3 m^{-3}), the combined volume fraction of soil organic matter and root biomass water equivalent ($SOM + B_R$ in m^3 m^{-3}). In order to calibrate the sensor one first has to determine the depth- (and distance-) weighted averages for ρ_{bd}, W_L, $SOM + B_R$ and θ as well as N_{pih} (averaged over 12 h) for the time of calibration. This is necessary because several factors can influence the ~~critical~~effective measurement depth z^* (which is the depth of the soil layer up to which 86 % of the neutrons that the CRS detects originate from) and the footprint size of the sensor (~~Fig. 4~~Fig. 2). Afterwards N_0 is adjusted iteratively (e.g. with a simple Solver routine in Microsoft Excel) until the right-hand side of the equation equals the left-hand side.

Kommentar [A38]: Rev1: The methods to determine soil organic carbon and root biomass water equivalents are not presented.

Kommentar [A39]: Reply: A description of the methods has been added. (See lines 190-195.)

We tested four soil moisture weighting approaches (Table ~~1~~2), described in detail below, to determine which information is necessary for an accurate calibration. ~~In a fifth approach we also tested whether including the influence of above-ground biomass (B_{ag}) further improves the performance of soil moisture retrieval with the CRS.~~

Kommentar [A40]: Rev1: This statement is too vague. I think what you meant here is that the objective performance measure is minimized, right?

Kommentar [A41]: Reply: True. The sentence was removed since we decided to omit the last approach.

1. In the first approach (simple depth-weighting, SDW) a linear depth-weighting function was used (Franz et al., 2012b), where $wt(z)$ represents the weight that is applied to the soil moisture measurements from a certain soil depth z:

$$\begin{cases} wt(z) = a\left[1 - \left(\frac{z}{z^*}\right)^b\right] & 0 \le z \le z^* \\ wt(z) = 0 & z > z^* \end{cases} \qquad (5),$$

where

$$a = \frac{1}{z^* - \frac{{z^*}^{b+1}}{(b+1){z^*}^b}} \qquad (6),$$

and

$$z^* = \frac{5.8}{\cancel{\frac{H_p}{\rho_w}} H_p + 0.0829} \qquad (7),$$

and

$$H_p = W_L + SOM + B_R + \cancel{\rho_w}\theta \qquad (8).$$

In these equations z is the soil depth below the surface in cm and z^* is the ~~critical~~effective measurement ~~soil~~ depth in cm, a is a parameter that ensures that the weights are conserved, b controls the curvature of the weighting function and equals~~is~~ 1 for linear weighting, ~~ρ_w is the density of water (here assumed to be 1 g cm^{-3}),~~ H_p is the water equivalent~~hydrogen content~~ of the belowground hydrogen pools (m^3~~g~~ m^{-3}), W_L is lattice water (m^3~~g~~ m^{-3}), SOM is soil organic matter water equivalent (m^3~~g~~ m^{-3}), B_R is root biomass water equivalent (m^3~~g~~ m^{-3}) and θ is the gravimetrically determined volumetric soil pore water content (m^3 m^{-3}). The original approach by Franz et al. (2012b) was modified by Bogena et al. (2013) using the total hydrogen content of belowground hydrogen pools H_p instead of just using the volumetric soil water content θ. Since H_p changes with soil depth we used an iterative approach to determine the appropriate weights. Starting with an average value for the upper 30 cm of the soil we computed an ~~critical~~effective measurement depth z^* and weighted H_p of the different soil depths accordingly. With this new value of H_p we then recomputed z^* and the weights. Usually the value of H_p stabilizes after a few iterations. The bulk density (ρ_{bd}) of the soil changes with depth and influences the soil moisture measurements too. Therefore it was also being taken into account during the iterative process of determining the ~~critical~~effective measurement depth z^* and the weighted soil moisture. In this first weighting approach we did not use our depth-specific measurements of W_L and $SOM+B_R$, instead we assumed an average weight fraction value of combined $W_L+SOM+B_R$ for the entire 30 cm profile.

2. The second approach (depth-specific weighting, DSW) was identical to the first one (SDW) except for using depth-specific measurements of W_L and SOM~~–~~+~~–~~B_R (see Table ~~2~~3 for an example).

3. For the third approach (distance-depth-weighting, DDW), we adopted the weighting approach described in Köhli et al. (2015). This approach introduces distance-dependent variable depth-weighting where the ~~critical~~effective measurement depth decreases with distance from the sensor. The ~~critical~~effective measurement depth z^* is calculated according to:

$$z^* = \rho_{bd}^{-1}\left[8.32 + 0.14 * \left(0.97 + e^{\frac{-r}{100}}\right) * \frac{26.42+H_p}{0.057+H_p}\right] \qquad (9),$$

where ρ_{bd} is the bulk density of the soil (g cm^{-3}), r is the radial distance (in meters) from the CRS and H_p is the water equivalent of the belowground hydrogen pools (m^3 m^{-3}) ~~total hydrogen content of belowground hydrogen pools~~ (see Eq. 8). This approach also assumes that the footprint size of the sensor varies with soil water content and atmospheric water content. We computed the varying footprint diameter for each calibration campaign and weighted the samples from 25, 75 and 200 m accordingly.

4. The fourth approach (distance-depth-weighting, non-linear, DDWnl) was identical to the third one (DDW) except for using the non-linear depth-weighting function recommend by Köhli et al. (2015) instead of the linear one (from Eq. 5):

$$wt(z) = e^{\frac{-2z}{z^*}} \qquad (10).$$

~~5. In the fifth approach, an above-ground biomass correction (ABC) was added to the third approach (DDW). This approach differs from the first four weighting approaches by explicitly correcting the neutron counts for vegetation effects, i.e., it corrects neutron counts for the additional damping by above-ground biomass without altering the depth weighting of the calibration function itself. To this end, we adopted the method proposed by Baatz et al. (2015) to~~

Kommentar [A42]: We removed this approach since the proposed above-ground biomass correction was developed for larger differences in biomass than our field site is subjected to.

~~further correct the neutron signal already corrected for pressure, incoming flux and water vapor (N_{pih}) and derive a vegetation corrected neutron count (N_{pihv}). According to Baatz et al. (2015) vegetation causes a neutron intensity reduction by 0.9% per kg of dry aboveground biomass (B_{ag}) per m^{2}:~~

~~$N_{pihv} = \frac{N_{pih}}{1-(0.009*B_{ag})}$ (11).~~

~~At our field site this means an intensity reduction of 57.3 % due to the beech forest surrounding the CRS (B_{ag} = 63.8 kg m^{-2}, see Sect. 3.2). The seasonal variation due to the presence of leaves on the trees is negligible (winter: B_{ag} = 62.8 kg m^{-2}; intensity reduction = 56.5 %), not even considering the fact that the leaves are still present as litter on the ground.~~

3.2. Estimation of biomass and influence of seasonal changes in biomass

Biomass influences neutron counts due to its hydrogen content. In order to test (and potentially exclude) the influence of seasonal changes in aboveground forest biomass, ~~a survey of the beech tree stand around the CRS was conducted. W~~we estimated living tree biomass and tree biomass changes throughout the year by applying the aboveground dry biomass functions for beech forest (*Fagus sylvatica* L.) from Santa Regina et al. (1997):

$$B_S = 0.0894 * DBH^{2.4679}$$ (~~12~~11),

$$B_B = 0.0317 * DBH^{2.3931}$$ (~~13~~12),

$$B_L = 0.0145 * DBH^{1.9531}$$ (~~14~~13).

B_S is dry stem biomass (kg tree^{-1}), B_B dry branch biomass (kg tree^{-1}), B_L dry leaf biomass (kg tree^{-1}) and *DBH* is the diameter of the tree stem at breast height (cm). Total dry above-ground biomass B_{ag} is the sum of the three components.

To apply these functions we conducted a survey of tree diameters and tree density in the beech forest that surrounds the CRS. This allowed us to determine both the total biomass of the beech forest, as well as the seasonally variable fraction of biomass (leaf biomass divided by total biomass). ~~The seasonally variable fraction of hydrogen mass in the trees aboveground can~~

~~introduce a second temporally dynamic signal on neutron counts. In order to determine this fraction w~~We first calculated the water mass (W_{agb}) in stems, branches and leaves (assuming a leaf water content of 0.6 kg per kg~~$^{-1}$~~ of wet biomass (Gravano et al., 1999) and a wood water content of 0.11 kg kg^{-1} (Bouriaud et al., 2004)). Finally, using the mass fraction of hydrogen in water (M_w = 0.1119 kg H per kg^{-}~~1~~ H_2O) and in dry biomass (M_b = 0.0622 kg H per kg~~$^{-1}$~~ Cellulose: $C_6H_{10}O_5$) ~~one can calculate~~ the total h~~ydrogen density~~ydrogen mass (H_{agb}) of above-ground biomass in the beech stand was derived:

$$H_{agb} = W_{agb} * M_w + B_{ag} * M_b$$ (~~15~~14).

We did not conduct surveys on the other tree species. Table 1 shows that the beech stand covers 56% of the footprint area around the CRS (when assuming the exponential distance-weighting from Zreda et al. (2008)). Pine covers 16%, spruce 13%, oak 8%. With the new distance weighting function of Köhli et al. (2015), the cover fractions of the other tree species would decrease even further. Also, the seasonal variation in spruce and pine above-ground biomass is very small and thus we consider it to be constant in this study.

~~and then converted the water and dry biomass values into hydrogen equivalents by assuming that the weight fraction of hydrogen in water is 0.1198 kg kg^{-1} (H_2O) and the hydrogen content in biomass is 0.0622 kg kg^{-1} (Cellulose: $C_6H_{10}O_5$).~~

~~The tree survey revealed a median diameter of 23.9 cm (Min: 3.2 cm, Q_{25}: 11.5 cm, Q_{75}: 43.7 cm, Max: 93.3 cm) and a tree density of 0.05 trees m^{-2}. With these values at hand and Eqs. (12)-(14) the dry above-ground biomass of the beech stand (B_{ag}) was computed to be 63.8 kg m^{-2} (with 62.8 kg m^{-2} from stem and branches and 1.0 kg m^{-2} from leaves) (Fig. 5). Assuming a water content of 0.11 kg kg^{-1} for wood and a water content of 0.6 kg kg^{-1} for leaves results in 9.2 kg m^{-2} of biomass water (W_{agb}) (with 7.8 kg m^{-2} from stem and branches and 1.5 kg m^{-2} from leaves). Finally, using the mass fraction of hydrogen in water (M_w = 0.1119 kg kg^{-1}) and in dry biomass (M_b = 0.0622 kg kg^{-1}) one can calculate the total hydrogen density (H_{agb}) of above-ground biomass in the beech stand:~~

~~$H_{agb} = W_{agb} * M_w + B_{ag} * M_b$ (15)~~

Kommentar [A43]: Rev3: is this 0.6 kg of H2O per kg of dry biomass? or wet biomass? clarify.

Kommentar [A44]: Rev3: this seems low. does this figure include free water (H2O in xylem) and cellulose-bound water (OH group)?

Kommentar [A45]: Reply: No, this just includes free water. The cellulose-bound water is calculated in the next step (assuming hydrogen content in dry biomass is 0.0622 kg kg-1).

~~Our calculations yielded a hydrogen density of 4.8 kg m^{-2} for stem and branches and a hydrogen density of 0.2 kg m^{-2} for leaves. Assuming that the hydrogen content of the stem and branches is constant and only the leaves change seasonally one is left with a fraction of variable hydrogen in the above ground biomass that accounts for 7.7 % of the total hydrogen mass. At high soil moisture, a 0.01 m^3 m^{-3} soil moisture change from 0.19 to 0.20 m^3 m^{-3} equals a change of 0.07 kg m^{-2} of hydrogen in the soil. At low soil moisture the change from 0.05 to 0.06 m^3 m^{-3} is equal to a change in hydrogen of 0.25 kg m^{-2} (due to the fact that the CRS also receives the neutron signal from deeper soil depths (larger critical depth z^*)).~~

Kommentar [A46]: Reply: Parts were moved to the results section (chapter 4.7).

3.3. Validation

Kommentar [A47]: Rev3: This section should provide more information, not just the equation. What is KGE? How is it used? What is compared with what? How are these variables computed (eg, r)? What is the significance of the result? Etc...

As an objective performance measure to compare the soil moisture time series derived from the CRS with the soil moisture time series from the TDT sensors we used the modified Kling-Gupta efficiency KGE' (Gupta et al., 2009; Kling et al., 2012):

$$KGE' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \qquad (\cancel{16}15).\cancel{,}$$

With correlation coefficient r~~,~~:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} * \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (16),$$

~~-~~bias ratio $\beta = \mu_{mod}/\mu_{obs}$ and variability ratio $\gamma = (\sigma_{mod}/\mu_{mod})/(\sigma_{obs}/\mu_{obs})$. The KGE' measures the Euclidian distance in a 3-D space where the correlation coefficient r is on one axis, the variability ratio β is on the second axis and the bias ratio γ is on the third axis. KGE' scores range from 1 (representing a perfect fit) to -∞. Due to the composite nature of the KGE' it is relatively simple to analyze which feature of the time series (correlation, bias, variability) contributes most to the good/bad performance of a model.

4. Results

4.1. Gravimetric soil water measurements and soil physical characteristics

Soil water content in the sandy soils ranged between 0.03 and 0.37 $m^3\ m^{-3}$ (absolute minimum and maximum values of individual soil core samples during the 10 sampling campaigns). The spatial distribution of volumetric soil water content for the 10 calibration days is shown in ~~Fig. 6~~Fig. 3. At each location the soil water content is an unweighted average value of the six samples taken from 0 to 30 cm depth. The mean volumetric soil water content for the calibration days over all calibration locations ranged from 0.07 up to 0.16 $m^3\ m^{-3}$ with standard deviations ranging from 0.015 to 0.047 $m^3\ m^{-3}$. The depth and distance weighted averages used for calibration ranged from 0.08 to 0.24 $m^3\ m^{-3}$ (see for example Table 4, column: θ_{depthW}). A general soil moisture pattern emerged with the soil moisture under coniferous tree stands being lower and under deciduous tree stands being higher. Especially the uppermost soil layer (0-5 cm) was drier under the coniferous trees – on average about 0.065 $m^3\ m^{-3}$ – while the deeper soil layers under coniferous trees were about 0.023 $m^3\ m^{-3}$ drier. The highest spatial variabilities in soil moisture were encountered during spring and fall seasons and more homogenous soil moisture conditions during winter and summer. The wettest calibration we conducted (WI~~SU~~) yielded an average soil water content of 0.29 $m^3\ m^{-3}$ for the top 5 cm. Calibration at higher soil water content is difficult as it only occurs for short periods of time after large precipitation events when significant amounts of intercepted water are also present in the canopy and litter layer.

The average bulk density (ρ_{bd}) measurements for the 10 calibration campaigns ranged from 1.16 to 1.22 $g\ cm^{-3}$ (mean: 1.18 $g\ cm^{-3}$, standard deviation: 0.02 $g\ cm^{-3}$). The weight fraction of soil organic matter and root biomass water equivalent ($wSOM + B_{\mathrm{R}}$) was determined to be 51.4 $g\ kg^{-1}$ in the shallowest soil layer (0-5 cm) with decreasing values at depth. The weight fraction of lattice water (wW_{L}) was determined to be 3.2 $g\ kg^{-1}$ in the shallowest soil layer with slightly increasing values at deeper soil depths.

4.2. Footprint variability

The footprint diameters calculated according to Köhli et al. (2015) and used in approaches 3 and 4 ranged from 185 m for the wettest to 200 m for the driest conditions. This resulted in distance weights of ~0.56 (for samples from 25 m distance), ~0.35 (for samples from 75 m distance) and ~0.10 (for samples from 200 m distance). These weighting factors varied only marginally between the individual calibration campaigns despite considerable differences in soil and atmospheric water content. Sampling distances with equal weights according to Köhli et al. (2015) would have differed from our sampling pattern (~1 m, ~33 m, ~140 m instead of 25 m, 75 m, 200 m), a condition which we balance by adjusting the distance weights. Furthermore the conditions within 30 m around our CRS are quite homogenous since the sensor is located within a pure beech stand and we are expecting little difference in average soil moisture content between locations at 1 and 25 m distance.

~~4.2.~~4.3. Calibration

The average reference atmospheric pressure (P_0) for the entire measurement period was 1005.8 hPa; the average reference incoming neutron flux (N_{avg}) was 328.3 ~~n h^{-1}~~ counts h^{-1}; the average reference absolute humidity (p_{v0}^{ref}) was 9.1 g m^{-3}. Equations (5) through (10) were used to calculate ~~the~~ depth-weighted volumetric soil water content ~~(~~θ_{depthW} and depth-weighted water equivalent of belowground hydrogen pools $(H_p)_{depthW}$~~)~~ according to the four weighting approaches we applied. Equations (1)-(3) were used to compute N_p, N_{pi} and N_{pih} ~~(as well as Eq. 11 to compute N_{pihv})~~, and then Eq. (4) to identify N_0 for each calibration. Table ~~2~~3 provides an example of the depth-weighting following approach 2 (DSW with depth-specific values of W_L and $SOM+B_R$).

The values in Table 3~~2~~ result in a depth-weighted average volumetric water content θ_{depthW} of 0.150 m^3 m^{-3}~~3~~, a depth-weighted water equivalent of belowground hydrogen pools~~volumetric water content including W_L and $SOM + B_R$~~ $(H_p)_{depthW}$ of 0.179 m^3 m^{-3} and a depth-weighted bulk density $(\rho_{bd})_{depthW}$ of 0.981 g cm^{-3}. If W_L and $SOM+B_R$ were not considered, the values for θ_{depthW} and $(\rho_{bd})_{depthW}$ would change to 0.146 m^3 m^{-3} and 1.013 g cm^{-3} respectively, because the ~~critical~~effective measurement depth z^* increases when the higher amounts of $SOM+B_R$ in the shallow layers are not considered, thus giving more weight to low soil moisture values in deeper soil horizons.

Kommentar [A48]: Rev1: The correct unit for incoming neutrons is “counts/sec”

Kommentar [A49]: Reply: Corrected.

~~4.3. Footprint variability~~

~~The footprint diameters calculated according to Köhli et al. (2015) and used in approaches 3-5 ranged from 185 to 200 m. This resulted in distance weights of ~0.56 (for samples from 25 m), ~0.35 (for samples from 75 m) and ~0.10 (for samples from 200 m). These weighting factors varied only marginally between the individual calibration campaigns.~~

Table 4~~3~~ lists the parameters relevant for calibration for all 10 calibration dates (again following approach 2, ~~S~~DSW, with depth-specific values of W_L and SOM-+-B_R).

Following the standard N_0-calibration approach of Desilets et al. (2010), we should have ended up with the same N_0 value for each of the 10 calibrations. However, the N_0 range we found was considerable – e.g. from 808~~58~~ to 895~~910~~ counts~~n~~ h^{-1} for the DDW approach (mean: 841.9~~78~~ ~~n h^{-1}~~counts h^{-1}, standard deviation: 13.7~~2113.~~5~~8~~ ~~n h^{-1}~~counts h^{-1}). As a consequence, the 10 computed time series ~~following Desilets et al. (2010)~~ also showed differences ~~of more than 0.1 m^3 m^{-3}~~ in volumetric soil water content~~(Fig. 7~~ (Fig. 4 illustrates results for the DDW approach).~~),~~ In the most extreme case, these differences were larger than 0.1 m^3 m^{-3} (which is equal to 24~~30~~ % of the total range of soil water content at the site)~~especially during conditions of high soil moisture (Fig. 7)~~.

In fact, none of the four~~ive~~ weighting approaches was able to solve the problem of determining a unique calibration parameter for our field site. All weighting approaches resulted in largely deviating N_0-values between the individual calibrations (see means and standard deviations in column 1 and 2 of Table 5). This in turn led to differences in the~~and~~ resulting time series of volumetric soil water content ~~between the individual calibrations~~ (see means and standard deviations in column 3 and 4 of Table ~~4~~5).

4.4. Modified~~New~~ calibration function

To include all information of our 10 calibration campaigns into our analysis, we fitted modified~~new~~ calibration functions to ~~our five~~four sets of 10 calibration points ~~each~~derived from the four~~4~~ different weighting approaches (see section 3.1). This was done by using the Microsoft Excel Solver software to optimize the three shape parameters (a_0, a_1, a_2) and N_0 through the

Kommentar [A50]: Rev1: The correct unit for the CRS measured neutron intensity is "counts/h"

Kommentar [A51]: Reply: Corrected throughout the revised manuscript.

Kommentar [A52]: Rev3: this is inaccurate; one time series shown in Fig. 7 has differences of that magnitude, whereas the other nine have considerably smaller deviations; please, state this result correctly, without creating undue alarm.

Kommentar [A53]: Reply: You are right in that our formulation is unclear and could be interpreted in different ways. So we modified the statement: 'As a consequence, the 10 computed time series following Desilets et al. (2010) also showed differences in volumetric soil water content (Fig. 4 illustrates results for the DDW approach). In the most extreme case, these differences were larger than 0.1 m^3 m^{-3} (which is equal to 30% of the total range of soil water content at the site).'.

Kommentar [A54]: Rev3: It is unclear how this conclusion was reached. Please describe in detail what was done, what were the results and how to interpret them.

Kommentar [A55]: Reply: We conclude that there is no unique calibration parameter and describe the calibration procedure in detail in the methods section (lines 256-303). The results are presented in Table 5 and Fig. 4 & 5. We added a more detailed interpretation of the results to the discussion section.

Kommentar [A56]: Rev1: According to Zreda et al. (2012) the presence of other hydrogen pools than soil moisture increases the stopping power of the soil, which leads to a change in the slope of the calibration function. Thus, calibration has to be performed using the total hydrogen pool, and soil moisture is then computed by subtracting other hydrogen pools than soil moisture from the measured neutron-derived soil moisture. It is unclear whether this procedure was applied in this study. If not, this would partly explain the differences in soil moisture estimates.

Kommentar [A57]: Reply: As outlined in the description of our four approaches (esp. Lines 267-285), we used the total hydrogen pool for calibration before subtracting other than soil moisture contributions according to Eq. 4 (Line 240).

Kommentar [A58]: Rev1: The term "new calibration function" is misleading. Changing the "a" parameters of the N0 calibration function is not new and was already presented by Iwema et al. (2015). They called this more adequately "modified N0 method". However, they only ...

Kommentar [A59]: Reply: You are right that 'new calibration function' is misleading. We changed all occurrences to 'modified calibration function'. It is true that the N0 parameter has a very similar influence on the shape of the calibration function as the a0 parameter. Still ...

calibration point cloud (solid lines in ~~Fig. 8~~Fig. 5). Plotting the N_{pih}-values of all 10 calibrations against the gravimetrically determined and depth- (and distance-) weighted volumetric soil moisture revealed that the standard shape of the soil moisture-neutron count relation is not valid at our field site. Instead of plotting along functions defined by the standard calibration (Desilets et al., 2010) (examples are dotted lines in ~~Fig. 8~~Fig. 5) our calibration points are better captured by less steep functions (solid lines in ~~Fig. 8~~Fig. 5 are the best-fit calibration functions for the different approaches). Using the N_0-calibration function with the standard shape parameters may lead to large soil water content deviations between individual calibration campaigns, especially under wet soil moisture conditions. The slope of the N_0-calibration function is essentially too steep, which means that in our environment a change in the neutron count is caused by a more subtle change in soil moisture than is assumed by the standard relationship – essentially the sensor has a higher resolution/sensitivity~~is more sensitive~~ than one would expect~~ed~~.

Kommentar [A60]: Rev1: Do you have any idea why?

Kommentar [A61]: Reply: Yes. If you look at Fig. 5 you see that at higher water content a smaller change in neutron counts is associated with a larger change in soil water content (the function is steeper). Therefore the uncertainty during the calibration is also larger.

Kommentar [A62]: Rev3: There is a lot of noise in Figure 5 and the difference between standard N0 and the improved data is not clear. Can you add error bars to the data points?

Kommentar [A63]: Reply: We are not sure about this comment. In our view, this difference between the standard and the modified calibration functions is clearly represented by the dotted and the solid lines respectively. Since there is already a lot going on in this figure, we decided not to add error bars and we were not sure of the added value they would provide.

The optimized parameters for the ~~five~~four approaches are shown in Table ~~5~~6. The resulting soil moisture time series are shown in ~~Fig. 9~~Fig. 6.

4.5. Validation

We tested whether the modified~~new~~ calibration functions improved the performance of the CRS measurements relative to in situ measurements, and if so, which of the weighting approaches performed best. In order to do that we compared the soil moisture time series from the CRS (using the standard N_0-calibration function from Desilets et al. (2010) and applying our newly derived corrected relationships) with the soil moisture time series from the TDT sensors distributed throughout the footprint. As a first step, the CRS measurements had to be converted to a soil water content value representative of the top 15 cm of the soil (the integration depth of the TDT sensors). For this purpose we compared the weighted volumetric water content (θ_{depthW}) from the gravimetric measurements of the calibration campaigns (basically what the CRS is supposed to "see") with the unweighted average gravimetric measurements of the top 15 cm (θ_{15cm}) (Fig. S2). We found strong linear correlations for two of the weighting approaches (SDW and DSW) with CRS water content being larger than the θ_{15cm} values and increasing differences for wetter soil conditions (indicating that for higher soil moisture the CRS overestimates soil water contents in the top 15 cm while for lower soil moisture the overestimation decreases). For

Kommentar [A64]: Rev1: Please provide a figure showing the comparison.

Kommentar [A65]: Reply: We have prepared a figure and will add it as a supplement (Fig S2).

approaches 3 and 4~~ 5~~ (DDW and~~,~~ DDWnl~~ and ABC~~) ~~an ~~offsets of 0.006~~5~~ and 0.011 m^3 m^{-3} indicated slightly lower weighted~~CRS~~ soil water content than the unweighted top 15 cm values. The linear correlations for the first two weighting approaches were expected since when it is wetter, the effective measurement~~penetration~~ depth is reduced for the CRS measurements and the wetter shallower soil layers receive more weight. Therefore, the CRS measurements result in higher soil water content than the gravimetric measurements. However, it seems that in approaches 3 and 4 the distance weighting counters this effect. A probable explanation is that the formula used for the distance-depth weighting increases the effective measurement depth. This causes higher weights for deeper (drier) soil layers even under wet conditions and could counteract the trend. We then converted the CRS time series by the above relationships into time series that were representative of the top 15 cm and compared them to the TDT measurements. The modified Kling-Gupta efficiency (KGE') was used as a performance measure. The worst performance was achieved by the simple depth weighting approach (KGE'(SDW) = 0.83, Table 7~~6~~), the performance improved when depth-specific weighting was included (KGE'(DSW) = 0.88) and it further improved when including distance weighting (KGE'(DDW) = 0.89~~2~~). The linear depth weighting worked better than the non-linear depth weighting (KGE'(DDWnl) = 0.83~~7~~). ~~The inclusion of a vegetation correction did not improve the performance any further (KGE'(ABC) = 0.92).~~ That means that the distance-depth-weighting approach (DDW) improved the neutron sensors performance the most. In comparison, using the single-point standard N_0-calibration function and DDW yielded KGE's for the individual calibration campaigns ranging from 0.58~~46~~ to 0.83~~79~~ with a mean KGE' of 0.71~~68~~ (±0.08~~9~~). It is important to note that all of the modified~~new~~ calibration approaches performed better than their standard calibration counterparts. The improvement of performance of the new N_0-calibration functions compared to the standard calibration functions was caused by the better agreement of both the bias ratios β and the variability ratios γ, i.e. both the means and the variabilities of the CRS time series better matched the TDT observations (see also ~~Fig. 10~~Fig. 7). This supports the hypothesis that at our field site larger than expected changes in neutron count are already caused by ~~more ~~subtle changes in soil moisture~~ than expected~~.

Kommentar [A66]: Rev1: Shouldn't the relationships vary with soil moisture content due changing sensor penetration depths?

Kommentar [A67]: Reply: Yes, that is what we expected and that is also what we found for the first two approaches. When it is wetter, the penetration depth is reduced for the CRS measurements and the wetter shallower layers receive more weight. Therefore, the CRS measurements show higher SWC than the gravimetrically determined SWC. However, it seems that the distance weighting counters this effect. A probable explanation is that the formula used for the distance-depth weighting increases the critical depth. This causes higher weights for deeper (drier) soil layers even under wet conditions and could counteract the trend.

Kommentar [A68]: Rev1: This finding is quite obvious given the insignificant changes in above biomass. Generally, the application of the vegetation correction makes only sense, when temporal biomass dynamics are expected and temporal information on biomass changes are available.

Kommentar [A69]: Reply: We removed the whole part on vegetation correction.

4.6. Optimizing calibration efforts

Kommentar [A70]: Rev1: This investigation is very similar to Iwema et al. (2015). Please discuss your results in the light of this study.

Kommentar [A71]: Reply: We added a discussion in section 5.1 (lines 687-695).

We further tested whether two or more individual calibration campaigns are required to determine a comprehensive calibration function shape, and under which soil moisture conditions these calibrations should be conducted. We paired each individual calibration point (derived from the best-performing weighting approach, DDW) with all the other calibration points (WI and S1, WI and S2, WI and S3, etc.) and computed best-fit calibration functions for all of these pairings (~~Fig. 11~~Fig. 8).

Then we used the resulting calibration functions to convert the measured neutron counts into time series of volumetric soil water content and compared these to the TDT measurements (again using the KGE' as the performance measure). We found that a two-point calibration proved to be sufficient in case that the difference in soil water content between the two calibrations was larger than~~at least~~ 0.1~~2~~ m^3 m^{-3} (i.e. for our sandy soils it covered ~50 % of the observed range of average soil water content). Figure 9 indicates that the calibrated neutron count-soil water content conversion will always perform well if the soil moisture difference between the two calibrations is sufficiently large. Also, it turned out to be more important to capture a calibration point at very dry rather than at very wet soil water contents. This is illustrated in ~~Fig. 12~~Fig. 9 where predominantly calibrations that involve low soil water contents (red dots) as the minimum value achieve KGE's of 0.9 while these KGE' values are also achieved more frequently with intermediate soil water contents (light blue dots) as the maximum value.

Kommentar [A72]: Rev1: The results plotted in Fig. 9 show clearly, that only the most extreme dry and wet samplings result in an acceptable calibration result, whereas sampling at intermediate soil moisture will lead to very uncertain calibration of the modified N0-method. On the other hand, this illustrates the value of the standard N0-method that will also produce stable results in case only one sampling date is available. Please add this to the discussion.

Kommentar [A73]: Reply: Fig. 9 shows that the best 2-point-calibrations are achieved with one sampling point taken under very dry conditions and another sampling point taken either under intermediate or wet conditions. In our case it is hard to see the value of the standard N0-method since it always resulted in too much soil moisture variability no matter whether the calibration was performed during wet, intermediate or dry conditions (because the standard calibration of N0 does not allow a change of the slope of the calibration function).

4.7. Variability of hydrogen pools~~Other potential influences on neutron count~~

Kommentar [A74]: Rev1: This chapter belongs to discussion

Kommentar [A75]: Reply: we moved parts of this chapter to discussion and only left the parts that really describe results. We also added results on other hydrogen pools, so we renamed the chapter

~~In search of potentially unaccounted factors that influence the neutron count we compared~~ ~~N_0-values obtained from the 10 calibrations with apparent atmospheric pressure, specific humidity, temperature and estimates of forest crown cover (derived from photographs taken from the ground aiming at the zenith) during the calibration campaigns. No seasonal or other temporal relationships were found.~~

The tree survey revealed a median diameter of 23.9 cm (Min: 3.2 cm, Q_{25}: 11.5 cm, Q_{75}: 43.7 cm, Max: 93.3 cm) and a tree density of 0.05 stems m^{-2}. With these values at hand and Eqs. (11)-(13) the dry above-ground biomass of the beech stand (B_{ag}) was 63.8 kg m^{-2} (with 62.8 kg m^{-2} from

Kommentar [A76]: Rev1: What about the other tree species?

Kommentar [A77]: Reply: We added to the text (Line: 342): 'We did not conduct surveys on the other tree species. Table 1 shows that the beech stand covers 56% of the footprint area around the CRS (when assuming the exponential distance-weighting from Zreda et al. (2008)). Pine covers 16%, spruce 13%, oak 8%. With the new distance weighting function of Köhli et al. (2015), the cover fractions of the other tree species will decrease even further. Also, the seasonal variation in spruce and pine above-ground biomass is very small and thus we consider it to be constant in this study.'

stem and branches and 1.0 kg m^{-2} from leaves) (Fig. 10). These values result in 9.2 kg m^{-2} of biomass water (W_{agb}) (with 7.8 kg m^{-2} from stem and branches and 1.5 kg m^{-2} from leaves). Further calculations yield a hydrogen mass of 4.8 kg m^{-2} for stem and branches and a hydrogen mass of 0.22 kg m^{-2} for leaves (Eq.14). Other hydrogen pools within the CRS footprint were also assessed. The thickness of the litter layer was determined to be 5 cm on average. Assuming a porosity of 85 % yields a hydrogen mass of 0.47 kg m^{-2} for a dry litter layer. Hence, the hydrogen mass of the static biomass (stem, branches and dry litter) amounted to 5.24 kg m^{-2}. Beech litter was found to have a maximum interception capacity of 2.8 mm in a forest in Luxembourg (Gerrits et al., 2010) corresponding to an additional 0.31 kg m^{-2} of hydrogen when the litter layer is wet. The canopy interception of beech can be assumed to be up to 1.5 mm (Gerrits et al., 2010) (i.e. another 0.17 kg m^{-2} of hydrogen is added to the system when the canopy is wet). The hydrogen contribution of soil organic matter and root biomass changes with soil water content because the effective measurement depth of the sensor changes. Applying the DDW approach we computed a value of 0.36 kg m^{-2} for wet conditions (0.29 m^3 m^{-3}), a value of 0.44 kg m^{-2} for intermediate conditions (0.17 m^3 m^{-3}) and a value of 0.66 kg m^{-2} for dry conditions (0.05 m^3 m^{-3}). The hydrogen contribution of lattice water also changes with moisture conditions (wet: 0.05 kg m^{-2}; intermediate: 0.07 kg m^{-2}; dry: 0.15 kg m^{-2}). A pore water content of 0.29 m^3 m^{-3} equals a hydrogen mass~~density~~ of 4.12 kg m^{-2}, a pore water content of 0.17 m^3 m^{-3} equals a hydrogen mass of 3.26 kg m^{-2} and a pore water content of 0.05 m^3 m^{-3} reduces the hydrogen mass to 1.77 kg m^{-2}. Figure 11 and Table 8 give an overview of the different hydrogen pools for varying moisture conditions within the footprint of the CRS.~~Assuming a linear depth-weighting function, the total amount of hydrogen from pore water that a CRS "sees" is 4.01 kg m^{-2} for a soil water content of 0.20 m^3 m^{-3} (critical depth = 17.9 cm) while it reduces to 3.93 kg m^{-2} for a soil water content of 0.19 m^3 m^{-3} (critical depth = 18.5 cm). That means that a change in volumetric soil water content of 0.01 m^3 m^{-3} is equal to a change in hydrogen of 0.08 kg m^{-2}. However, the same change in soil water content under drier conditions is associated with a larger change in hydrogen: if the soil water content is 0.06 m^3 m^{-3} (critical depth = 31.6 cm), the CRS "sees" 2.12 kg m^{-2} of hydrogen, if the soil water content is 0.05 m^3 m^{-3} (critical depth = 33.4 cm) then the CRS "sees" only 1.87 kg m^{-2} — so the difference in hydrogen is 0.25 kg m^{-2}. The variability in hydrogen due to foliation and defoliation in the beech forest surrounding the CRS amounts to 0.22 kg m^{-2}. This means that it equals a change in soil water content of about 0.031 m^3 m^{-3} (under wet conditions) and 0.009 m^3 m^{-3} (under dry conditions). These differences for wet and dry~~

~~conditions are due to the fact that the critical depth of the sensor is larger during dry conditions and therefore an equal increase in soil water content requires a larger amount of water since a larger soil column has to be filled. These calculations disregard the fact that fallen leaves still contain hydrogen (which hence is not completely removed from the system immediately and therefore should also reduce the expected variability). At our field site 65 % of the distance-weighted area surrounding the CRS is covered by deciduous trees (mainly beech and oak), the other 35 % do not experience a significant annual cycle of leaf growth and fall (pine, spruce and grassland). This should further reduce the influence of seasonally variable biomass on the cosmic ray neutron counts (with a potential maximum influence of leaf-out during wet conditions of 0.020 m^3 m^{-3} and only 0.006 m^3 m^{-3} in dry conditions). In summary, we do not expect a significant impact of seasonally varying above-ground biomass on the measurements of soil water content.~~

5. Discussion

5.1. Potential influences on neutron counts

The 10 N_0-calibration parameters derived from our 10 calibrations varied considerably. In a first analysis we found that this was not related to the different soil moisture conditions during calibration. In search of other potentially unaccounted factors that influence the neutron count we compared N_0-values obtained from the 10 calibrations with apparent atmospheric pressure, specific humidity, temperature and estimates of forest crown cover (derived from photographs taken from the ground aiming at the zenith) during the calibration campaigns. No seasonal or other temporal relationships were found. The contributions of different hydrogen pools (Fig 11) reveal that a large percentage of hydrogen at our field site stems from the above-ground vegetation (52 to 68 %, depending on moisture conditions). Fortunately, most of this hydrogen is static in nature and can be accounted for by the calibration of the CRS. Assuming that the hydrogen content of the stem and branches is constant and only the leaves change seasonally one is left with a fraction of variable hydrogen in the above-ground biomass that accounts for 2-3 % of the total hydrogen mass. The variability in hydrogen due to foliation and defoliation in the beech forest surrounding the CRS amounts to 0.22 kg m^{-2}. This means that it equals a change in

Kommentar [A78]: Rev1: So the whole discussion of this chapter is unimportant and should be reduced to 1-2 sentences.

Kommentar [A79]: Reply: Would you say that just because our results suggest that seasonally-varying above-ground biomass does not influence the neutron count significantly the discussion of this finding is not important? We think this finding is very important for the use of CRS in forested areas and worth the extended calculation and discussion. (In the end, it makes life much easier when applying CRS in forests).

Kommentar [A80]: We moved parts of this section to discussion (chapter 5.1).

Kommentar [A81]: Rev1: This chapter belongs to discussion

Kommentar [A82]: Reply: Moved.

Kommentar [A83]: Rev3: isn't there a seasonality or some other temporal variability (eg, with droughts) in the free water within trunks and branches?

Kommentar [A84]: Reply: Yes, there is some seasonality also in this regard (there is even daily fluctuations indicating variations in transpiration flux). This, however, was not part of our analysis since it is difficult to determine the exact numbers and it is likely that these variations are too small to influence the neutron count.

soil water content of about 0.031 ~~cm~~m^3 ~~cm~~m^{-3} (under wet conditions) and 0.018 ~~cm~~m^3 ~~cm~~m^{-3} (under dry conditions). These differences for wet and dry conditions are due to the fact that the ~~critical~~effective measurement depth z* of the CRS increases for dry conditions: the sensor receives the neutron signal from deeper soil depths and therefore an equal increase in soil water content requires a larger amount of water since a larger soil column has to be filled. At high soil moisture, a 0.01 m^3 m^{-3} soil moisture change from 0.28 to 0.29 m^3 m^{-3} equals a change of 0.07 kg m^{-2} of hydrogen in the soil. At low soil moisture the change from 0.05 to 0.06 m^3 m^{-3} is equal to a change in hydrogen of 0.12 kg m^{-2}. ~~The variability in hydrogen due to foliation and defoliation in the beech forest surrounding the CRS amounts to 0.22 kg m^{-2}. This means that it equals a change in soil water content of about 0.031 cm^3 cm^{-3} (under wet conditions) and 0.018 cm^3 cm^{-3} (under dry conditions).~~The above calculations with respect to biomass variability disregard the fact that fallen leaves still contain hydrogen (which hence is not completely removed from the system immediately and therefore should also reduce the expected variability). At our field site 65 % of the distance-weighted area surrounding the CRS is covered by deciduous trees (mainly beech and oak), the other 35 % do not experience a significant annual cycle of leaf growth and fall (pine, spruce and grassland). This should further reduce the influence of seasonally variable biomass on the cosmic ray neutron counts (with a potential maximum influence of leaf-out during wet conditions of 0.020 m^3 m^{-3} and only 0.012 m^3 m^{-3} in dry conditions). In summary, we do not expect a significant impact of seasonally varying above-ground biomass on the measurements of soil water content. Also, we could not find systematic changes in the calibration results connected to the annual cycle of tree foliation/defoliation (i.e. a reduction in counts during summer due to higher hydrogen content in the above-ground biomass). Therefore we deem a correction for variable hydrogen from forest canopy biomass~~ hydrogen~~ at different times of the year unnecessary.

With regard to other varying hydrogen pools we noticed that the influence of interception storage both in the canopy and in the litter layer can potentially have an impact. When both the canopy and the litter layer are wet, the combined hydrogen amount within these two stores can sum up to almost 5 % of the total hydrogen pool equaling a change in volumetric soil water content of 0.067 m^3 m^{-3} (Fig. 11). It is not possible to solve this problem by calibrating during conditions of high interception storage since then the soil water content would be underestimated as soon as the canopy is dry. Calibration during conditions of dry canopy and litter layer is recommendable

Kommentar [A85]: Rev1: This is only true when assuming that the CRS footprint is completely covered by beech, which is however not the case.

Kommentar [A86]: Reply: That is true. So we present an extreme case here since the other vegetation types experience smaller seasonal changes in above-ground biomass. In reality we should expect even less variation in neutron counts due to foliation/defoliation.

Kommentar [A87]: Rev1: So the whole discussion of this chapter is unimportant and should be reduced to 1-2 sentences.

Kommentar [A88]: Reply: Would you say that just because our results suggest that seasonally-varying above-ground biomass does not influence the neutron count significantly the discussion of this finding is not important? We think this finding is very important for the use of CRS in forested areas and worth the extended calculation and discussion. (In the end, it makes life much easier when applying CRS in forests).

because conditions with an empty interception store are generally prevalent and can be much better defined than conditions with a filled interception store. A potential solution to the influence of the variable interception storage filling is the introduction of another neutron count correction using observed, derived or modeled interception storage values (similar to the pressure or the water vapor correction).

~~The tenfold standard calibration of our CRS produced 10 different time series of volumetric water content. The differences between the individual time series at times exceeded 0.1 m^3 m^{-3}. Moreover, the time series of soil water content derived from the neutron counts via the standard N_0 calibration function exhibited a variability that was too high compared to the distributed continuous in situ measurements. Altering the shape of the calibration function led to much higher congruence between the individual calibration efforts. Furthermore, the determination of a new calibration function enhanced the performance of the CRS measurements significantly when comparing them with independent distributed measurements of soil water content. Different weighting approaches proved to be more or less useful in identifying appropriate soil water contents for the time of calibration campaigns.~~ The fact that the depth-specific weighting (DSW) approach performed better than the simple depth weighting (SDW) is an indication that the depth variations in lattice water, soil organic matter and root biomass content should be explicitly ~~represented~~ accounted for during the calibration of the CRS. The best performance was achieved with a weighting approach (DDW) that explicitly takes into account both depth-weighting as well as distance weighting of the soil water content (Table 7). This suggests that the variation in the footprint diameter needs to be considered during individual calibration campaigns. Linear depth-weighting resulted in a better CRS performance than non-linear depth-weighting since the non-linear depth-weighting basically underestimated soil water contents during wet periods (because higher weights of deeper (drier) soil layers were included). This caused both a decrease in the mean soil water content as well as a decrease in the variability of the soil water content time series and hence reduced the performance of the CRS. In soils where water content increases with depth the difference between linear and non-linear depth-weighting could be smaller (even negligible), at our field site, however, the decrease of water content with depth apparently favors~~requires~~ the use of a linear depth-weighting function.~~Adding a correction for above-ground biomass to the time series of neutron counts (converting N_{pih} to N_{pihv} using Eq. 11) did not improve the performance of the CRS measurements. It only marginally changed the shape of the~~

~~calibration function and produced almost the same time series of soil water content as the version without any correction for above ground biomass. Also, we could not find systematic changes in the calibration results connected to the annual cycle of tree foliation/defoliation (i.e. a reduction in counts during summer due to higher hydrogen content in the above ground biomass). Furthermore, our calculations of variable hydrogen mass in the canopy suggested that these seasonal changes are small compared to the changes of hydrogen mass in the soils caused by changes in soil water content. Therefore we deem a correction for variable forest canopy hydrogen at different times of the year unnecessary.~~

Kommentar [A89]: Reply: We restructured the discussion.

The differences in calibration results are ~~more~~ likely caused by the fact that the shape of the N_0-calibration function is different at our field site. That means that while being temporally stable the shape of the calibration function is spatially variable – there is no standard curve applicable to all sites. At our site the function is less steep than the standard N_0-calibration function suggested by Desilets et al. (2010), i.e. a similar increase in neutron counts is associated with a smaller decrease in soil moisture. A recalibration of the shape of the curve using all calibration points considerably improved the agreement between in situ measurements and CRS measurements of soil moisture. A two-point calibration already proved to be sufficient to define the correct shape of the calibration function given that the soil moisture states at the two calibration times we~~a~~re sufficiently different. In a recent study Iwema et al. (2015) also investigated temporal field sampling strategies for three different calibration methods. They tested combinations of different numbers of random sampling dates and found that using more than six random sampling dates did not improve their calibration results much more. However, for the N_0-calibration method they found that selecting sampling dates with distinct soil wetness conditions could reduce the required number of sampling~~ date~~s. In conclusion they also recommended more than one calibration campaign for the N_0-calibration approach and argued that the shape of the calibration function should not be fixed but kept variable during the calibration process. This is in line with our findings on the shape of the calibration function.

Kommentar [A90]: Rev1: Please discuss your results in the light of the results found by Iwema et al. (2015).

Kommentar [A91]: Reply: We added a discussion of the findings of Iwema et al. (2015).

We can only speculate about the reasons behind this shape inconsistency of the calibration function for our site since we did not do any theoretical neutron modeling. To our knowledge ~~at our site~~ we are dealing with the lowest number of counts of all published studies (average N_0 = 878 ~~n h^{-1}~~counts h^{-1}, Table 4~~3~~). Although the calibration function was theoretically developed for all environments it has not~~probably never~~ yet been tested sufficiently in such low-count, forested environments. Moreover, due to the low neutron count the uncertainty in the determination of soil water content during calibration has a much higher influence on the calibration results than in high-count environments. ~~And while the shape of the function seems to work well in high-count environments, it clearly does not at our site.~~ Bogena et al. (2013) pointed out another complicating factor that is present in forested environments – the litter layer. They showed that at their sites (N_0: 913 to 1397 ~~n h^{-1}~~counts h^{-1}) the model-derived water content within the litter layer (under spruce) was subject to much higher variability than the water content in the underlying soil. During wet conditions the water within the litter layer contained 36 % of the hydrogen mass within the footprint of the CRS while during dry conditions it contained only 10 % of the hydrogen mass. This leads to an increase in the variability of the neutron counts and can thus cause an overestimation of soil water content during wet conditions. Although the water within the litter layer at our site accounts for a much smaller fraction of the total hydrogen pool (up to 3 %) it can still have an influence on the neutron counts and the calibration results. The occurrence of canopy interception would have the same variability-increasing effect on the CRS signal, although it is expected to be significantly smaller than the influence of the litter layer. ~~We argue that an adjustment to the shape of the calibration function is able to solve this problem. By decreasing the slope of the calibration function we effectively reduce the sensitivity of the CRS and hence the temporal variability in the output signal (the time series of soil water content).~~ Baatz et al. (2014) working also in a low-count environment (N_0: 936 to 1242 ~~n h^{-1}~~counts h^{-1}) with land use ranging from grassland to agriculture to forest compared the standard N_0-calibration method to another calibration method developed by Shuttleworth et al. (2013) (the COSMIC operator) and found that the former interpreted dry periods drier and wet periods wetter – which is ~~also~~ in accordance to our findings that suggest that the standard N_0-calibration function is too steep. Lv et al. (2014), in a study at a mixed-forest/grassland site also recommended more than one calibration. They operated in a high-count environment in Utah (N_0 = 2189 ~~n h^{-1}~~counts h^{-1}) and attributed the different shape of their calibration function to binary soil moisture patterns at their site where the grassland soils were much drier than the forest soils under wet conditions but just as dry under dry conditions. Our field site is subject to similar spatial variability since it is also comprised of multiple areas with non-uniform soil water content (mean values of soil water contents differ between different forest stands). Following the argumentation of Lv et al. (2014), t~~T~~he fact that distance weighting improved our results can be regarded as an indication that non-

Kommentar [A92]: Rev2: Franz 2013 WRR investigated the impact of horizontal heterogeneity on the signal.

Kommentar [A93]: Reply: We are discussing the topic of horizontal heterogeneity in line 724-736. Here, we are just comparing the count rates of different studies (and in Franz et al. (2013) the lowest count rates are also above 1000 cph.

Kommentar [A94]: Rev3: The absolute count rate has no influence on the shape of the response function, just on the precision of calibration. I would remove this conclusion or reword it to make this conclusion ("it clearly does not at our site") less strong.

Kommentar [A95]: Reply: Ok. We removed this conclusion.

Kommentar [A96]: Rev3: but on page 9831 (of my copy) you stated that the sensitivity is better when using your calibration. Please, make these two statements consistent.

Kommentar [A97]: Reply: On page 9831 we stated that the sensitivity of the sensor is essentially higher than it should be (not better). This means that already a small difference in neutron counts indicates a large difference in soil moisture. The modified calibration accounts for this by decreasing the slope of the calibration function and thereby reducing the sensitivity of the sensor (so that now a bigger difference in neutron counts is required to cause differences in the soil water content reading). We modified our statement on page 9831 (= manuscript line 475) to: '…essentially the sensor has a higher resolution/sensitivity than one would expect.'

homogeneous soil moisture conditions indeed~~also~~ lead to changes in the shape of the calibration function. At our site, distance weighting reduced the spatial variability within the footprint of the sensor since it assigned higher weights to the closest sampling sites which were all located in the homogenous and relatively wet beech forest, while the influence of the drier soils under the coniferous trees was reduced. ~~In a recent study Iwema et al. (2015) investigated temporal field sampling strategies for three different calibration methods. They also recommend more than one calibration campaign for the N_0-calibration approach and argue that the shape of the calibration function should not be fixed but variable during the calibration process.~~

Kommentar [A98]: Rev1: This statement is not clear to me. Please explain in more detail.

If it was possible to fully correct for all factors that influence footprint size, depth-weighting and neutron count, a one-time calibration of the CRS would be sufficient. However, the abundance of different hydrogen pools and the uncertainties in the sensing depth estimation will always lead to uncertainties in the calibration process. Therefore we argue~~think~~ that for~~when intending to~~ the use of the CRS as a simple tool to measure soil water content at intermediate scales, the ~~benefit of obtaining all corrections does not justify the~~ efforts of~~required to~~ measuring~~e~~ all necessary parameters are not justified~~necessary~~. As shown by Iwema et al. (2015) and by the results of this study, this issue can be dealt with by using site-specific calibration parameters estimated from in-situ samples taken during dry and wet conditions. Hence,~~Therefore~~ we recommend a two~~-time-~~point calibration that – although being empirical in nature – inherently incorporates many of the required corrections.

Kommentar [A99]: Rev1: This part is somewhat misleading. Corrections of the neutron count rate (Eqs. 1-3) are essential for any application of the CRS (e.g. Zreda et al., 2012). Vegetation correction is only needed for sites with significant biomass changes. On the other hand, the characterization of the temporal stable hydrogen pools is important for the application of the N0-method. However, the abundance of different pools and the uncertainties in the sensing depth estimation will always lead to uncertainties in the calibration process. As shown by Iwema et al. (2015) and by the results found in this study, this issue can be partly circumvented by the using site specific calibration parameters estimated at using in-situ samples taken during dry and wet conditions. Please reformulate in this sense.

Kommentar [A100]: Reply: Changed to: 'If it was possible to fully correct for all factors that influence footprint size, depth-weighting and neutron count, a one-time calibration of the CRS would be sufficient. However, the abundance of different hydrogen pools and the uncertainties in the sensing depth estimation will always lead to uncertainties in the calibration process. Therefore we argue that for using the CRS as a simple tool to measure soil water content at intermediate scales, the efforts of measuring all necessary parameters are not justified. As shown by Iwema et al. (2015) and by the results of this study, this issue can be partly circumvented by using site-specific calibration parameters estimated from in-situ samples taken during dry and wet conditions. Hence, we recommend a two-time calibration that – although being empirical in nature – inherently incorporates many of the required corrections.'.

6. Conclusion

Our results suggest that a one-time calibration of the CRS using the available neutron count corrections and weighting approaches is not sufficient at our field site. This is mainly due to the fact that the shape of the standard N_0-calibration function is not able to reproduce~~capture~~ the dynamics in soil water content we observed with our network of distributed in situ TDT sensors. Several factors could cause this discrepancy, amongst them the presence of a litter layer~~s~~ and spatially heterogeneous soil moisture conditions within the sensor footprint. After calibrating the CRS 10 times in a mixed forest in north eastern Germany we found that a two-point calibration already considerably improved~~s~~ the agreement between soil water content derived from in situ

TDT measurements and from the CRS, given significantly different moisture conditions during the two calibration periods/campaigns (for a detailed explanation on the procedure see Appendix A). We found that the explicit consideration of depth-specific values of soil organic matter and root biomass improved the calibration results while ~~taking into account and~~ seasonal changes in above-ground biomass in the forest ~~was unnecessary~~were found to be negligible~~ because of their small amplitude~~. While there is no doubt that further investigations~~s~~ ~~on~~ of factors that influence the neutron signal are necessary and useful, it is also apparent that it becomes increasingly difficult to distinguish between the effects of the individual correction factors and the uncertainty caused by all the corrections. Therefore our goal was to use empirical data to test available methods and combinations thereof and to provide a guideline on how to easily and comprehensively calibrate a CRS in various environments using these methods. Looking beyond that objective, ~~investigations in the form of~~ site intercomparison studies along gradients from high to low-count environments and/or from locations with varying litter layers could give rise to the development of simple corrections to the shape of the N_0-calibration function.

Kommentar [A101]: Rev1: Actually, the seasonal changes of the hydrogen pools in this forest site are negligible. Thus vegetation correction can be omitted.

Kommentar [A102]: Reply: Agreed.

When measuring soil water content with a CRS it is important to note that over time the measurements are hardly ever representative of the exact same soil segment around and below ~~one~~ the sensor (Köhli et al, 2015). With the footprint shrinking and expanding and the ~~critical~~effective measurement depth in~~of~~ the soil decreasing and increasing we have to be careful when interpreting and using our results. If we keep that in mind, however, this new technology will indeed be able to bridge the gap between point in-situ and areal remote sensing soil moisture measurements and thus provide a valuable tool for the advancement of hydrologic understanding.

Kommentar [A103]: Rev1: This statement is based on Köhli et al. (2015), but not on results of this study and thus should be omitted.

Kommentar [A104]: Reply: Since this a very important statement and should be considered by everybody using a CRS, we would like to keep it. But we added the proper reference (Köhli et al.).

Appendix A: Best practice for calibration in low-count forest environments

We provide an Excel file as a supplement to perform the calculations described in the following step-by-step instructions.

1. Set up (or use) a weather station that monitors air temperature and relative humidity close to the CRS.

2. Set up the CRS in a location where the conditions within a radius of at least 30 m around the sensor are relatively homogeneous (similar soils, tree species, expected soil moisture conditions).
3. Switch on the CRS and come back later for calibration (or set it up before 6 a.m. and start calibrating on the same day). You should at least have 12 hours of CRS data for one calibration. Do not switch it off after the calibration, let it record continuously.
4. Choose a day with very dry or very wet soil moisture conditions for the first calibration campaign and wait for the opposite conditions for your second calibration (this might take a full year to achieve, but you will not lose any data, you will just not be able to accurately convert the data immediately).
5. Choose days without rain or snow for your calibrations, litter and canopy should be dry.
6. Take 108 soil samples from 18 locations (six directions, three distances) and six depths (0-30 cm). For equal distance weights choose distances according to Köhli et al. (2015) (~1, ~33 and ~140 m). ~~according to Franz et al. (2012b)~~
7. Weigh the samples the same day you take them, let them oven-dry for 24 h at 105°C and weigh them again to determine the volumetric water content (θ) and the bulk density (ρ_{bd}).
8. Create six bulk samples from the six different soil depths (2 g from each of the 18 locations suffices for each soil depth).
9. ~~Analyze~~ Determine the combined soil organic matter (SOM) and root biomass (B_R) content of the six bulk samples by weighing them (after regular oven-drying at 105°C) and then heating them to a temperature of 400°C for 24 h before weighing them again. Convert SOM and B_R to water equivalents by multiplying the~~ir~~ weight by 0.556.

   Caution: In clay-rich soils this method tends to overestimate soil organic matter content because some of the lattice water is removed already at temperatures around 400°C (Howard and Howard, 1990).
10. ~~Analyze~~ Determine the lattice water (W_L) content of the six bulk samples by weighing them (after SOM and B_R extraction at 400°C) and then heating them to a temperature of 1000°C for 24 h before weighing them again.

    Caution: Carbonate-rich soils experience thermal breakdown of carbonates at temperatures above 430°C (Ben-Dor and Banin, 1989).
11. ~~9.~~ Determine the water equivalent of the average hydrogen content of belowground hydrogen pools (H_p) for each soil depth.

Kommentar [A105]: Rev1: This step is obvious and should be omitted.

Kommentar [A106]: Reply: When we first set up the sensor and calibrated it for the first time we brought a battery with us to let the sensor run. After we had finished the soil collection, we took the battery back home with us. Only later it became clear that we should have collected neutron counts for a longer period of time. So we could not use the data from our first calibration effort. This was a hard lesson to learn and we want to make sure that other people do not make the same, admittedly stupid, mistake.

Kommentar [A107]: Rev1: The sampling locations should be adapted to the footprint estimates after Köhli et al. (2015).

Kommentar [A108]: Reply: We adapted the sampling distances to the footprint estimates after Köhli et al. (2015).

Kommentar [A109]: Rev3: Franz et al 2012 is not the correct reference; Franz et al merely repackaged the information given in Zreda et al 2012.

Kommentar [A110]: Reply: We decided to recommend sampling distances according to Köhli et al. (2015). Concerning the sampling pattern we used: Franz et al. 2012 describe 3 circles with distances around the CRS of 25, 75 and 200 m. Zreda et al. 2012 describe 3 circles with distances around the CRS of 25, 75 and 175 m. Since we used 25, 75 and 200 m the more correct citation to describe our calibration setup would be Franz et al. 2012. However, to recognize the contribution of Zreda et al. we inserted the reference in line 174: '…we followed the recommended sampling pattern for the calibration of CRS which was developed by Zreda et al. (2012) and slightly modified and detailed in Franz et al. (2012b).'.

Kommentar [A111]: Rev2: Again, is this method for lattice water supported by refs? If not then should be noted that this is a pragmatic procedure with expected minimal error for most soil groups other than volcanics, . . . etc. (?). Unfortunately I don't know all the soil groups this might be affected by so hopefully a pedologist can set us straight.

Kommentar [A112]: Reply: The method of heating the samples to a temperature of 1000°C to determine lattice water was used in many CRS studies (e.g. Zreda et al., 2012; Bogena et al., 2013) The only complication we found occurs in carbonate-rich soils where thermal breakdown of carbonates will contribute to the lattice water account. We added two cautions to the recommendations we give in the appendix.

Equation (8).

~~10.~~12. Apply a linear weighting function to your gravimetrically determined H_p measurements accounting for the change in the ~~critical~~effective measurement depth z* of the sensor and retrieve a weighted average of H_p within the footprint of the CRS by iteration. Start out by computing the ~~critical~~effective measurement depth z* corresponding to your gravimetrically determined values of H_p and ρ_{bd} averaged over the entire 30 cm. Then apply the weights for the different soil depths z and update the values. Recalculate the ~~critical~~effective measurement depth z* and continue this procedure until all values stabilize. Do this for each sampling/calibration distance (~1, ~33 and ~140 m~~25, 75 and 200 m~~) separately.

Equations (5), (6) and (9).

~~11.~~13. Apply an additional distance-weight to the depth-weighted volumetric water contents from the different locations in order to account for variations in the footprint size. Also do this iteratively adjusting H_p and the distance weights until both become stable.

Equations are conveniently provided as a supplement by Köhli et al. (2015) in the form of an Excel file~~sheet~~.

~~12.~~14. Use the depth-and-distance weights to compute weighted values of soil water content (θ), bulk density (ρ_{bd}), lattice water (W_L), soil organic matter and root biomass water equivalent (SOM+B_R).

~~13.~~15. Average raw neutron counts (N_{raw}) from the moderated sensor (measuring fast neutrons) over 12 h with a moving window.

~~14.~~16. Retrieve data from the neutron monitor close to your location in order to correct for the varying intensity of incoming neutrons (you may have to correct this data and fill gaps).

~~15.~~17. Using the entire time series for the period where cosmic-ray data is available determine average atmospheric pressure (P_0), average incoming neutron intensity (N_{avg}) and average absolute humidity (p_{v0}^{ref}).

~~16.~~18. Correct raw neutron counts for atmospheric pressure variations (N_p).

Equation (1).

~~17.~~19. Correct raw neutron counts for incoming neutron intensity variations (N_{pi}).

Equation (2).

~~18.~~20. Correct raw neutron counts for absolute humidity variations (N_{pih}).

Equation (3).

~~19. Plot the N_{pih} of both calibrations against the gravimetrically measured, distance- and depth-weighted volumetric soil water content (θ) according to the standard N_0-calibration function with fitting parameters.~~

~~Equation (4).~~

21. Fit a function through the two calibration points altering N_0, a_0, a_1 and a_2 (e.g. using Microsoft Excel solver). When doing this, use average values of the two calibration campaigns for bulk density (ρ_{bd}), lattice water (W_L), soil organic matter and root biomass water equivalent (SOM~~-~~+~~-~~B_R).

22. Plot the N_{pih} of both calibrations against the gravimetrically measured, distance- and depth-weighted volumetric soil water content (θ).

~~20.~~23. Use best fit parameters to convert time series of N_{pih} to volumetric soil water content.

Acknowledgement

Funding was provided by the Terrestrial Environmental Observatories (TERENO) and the Virtual Institute for Integrated Climate and Landscape Evolution (ICLEA). We would like to thank the Müritz National Park for allowing us to conduct~~site~~ our research in their forest. Marvin Reich, Iris Heine, Lisei Köhn, Janek Dreibrodt, Stephan Schröder, Erik Reinholz, Christian Rippich, Christopher Gravesen, Jörg Wummel all helped out in the field while Philip Müller and Hans-Peter Nabein assisted in the lab. Gabriele Baroni, Lena Scheiffele and Katja Mroos lent us their field equipment and Martin Schrön provided us with scripts for depth-distance-weighting. We thank Heye Bogena and two anonymous referees for their constructive feedback which helped us a lot to improve the manuscript.

References

Baatz, R., Bogena, H.R., Hendricks Franssen, H.-J., Huisman, J., Qu, W., Montzka, C., and Vereecken, H.: Calibration of a catchment scale cosmic-ray probe network: A comparison of three parameterization methods, J. Hydrol., 516, 231-244, doi:10.1016/j.jhydrol.2014.02.026, 2014.

Baatz, R., Bogena, H.R., Hendricks Franssen, H.-J., Huisman, J.A., Montzka, C., and Vereecken, H.: An empirical vegetation correction for soil water content quantification using cosmic ray probes, Water Resour. Res., 51, 2030–2046, doi:10.1002/2014WR016443, 2015.

Bachelet, F., Balata, P., Dyring, E., and Iucci, N.: Attenuation coefficients of the cosmic-ray nucleonic component in the lower atmosphere, Il Nuovo Cimento, 35, 23-35, doi:10.1007/BF02734822, 1965.

Ball, D.F.: Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils, J. Soil Sci., 15, 84-92, 1964.

Baroni, G. and Oswald, S.: A scaling approach for the assessment of biomass changes and rainfall interception using cosmic-ray neutron sensing, J. Hydrol., 525, 264-276, doi:10.1016/j.jhydrol.2015.03.053, 2015.

Ben-Dor, E., and Banin, A.: Determination of organic matter content in arid-zone soils using a simple “loss-on-ignition” method, Commun. Soil Sci. Plan., 20, 1675-1695, 1989.

Bogena, H.R., Huisman, J.A., Baatz, R., Hendricks Franssen, H.-J., and Vereecken, H.: Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: the worst case scenario, Water Resour. Res., 49, 5778–5791, doi:10.1002/wrcr.20463, 2013.

Bouriaud, O., Bréda, N., Moguédec, G., and Nepveu, G.: Modelling variability of wood density in beech as affected by ring age, radial growth and climate, Trees, 18, 264–276, doi:10.1007/s00468-003-0303-x, 2004.

Coopersmith, E., Cosh, M., and Daughtry, C.: Field-scale moisture estimates using COSMOS sensors: a validation study with temporary networks and Leaf-Area-Indices, J. Hydrol., 519, 637-643, doi:10.1016/j.jhydrol.2014.07.060, 2014.

Davies, B.E.: Loss-on-ignition as an estimate of soil organic matter, Soil Sci. Soc. Am. J. 38, 150-151, 1974.

Desilets, D. and Zreda, M.: Spatial and temporal distribution of secondary cosmic-ray nucleon intensities and applications to in situ cosmogenic dating, Earth Planet. Sc. Lett., 206, 21-42, doi:10.1016/S0012-821X(02)01088-9, 2003.

Desilets, D. and Zreda, M.: Footprint diameter for a cosmic-ray soil moisture probe: theory and Monte Carlo simulations, Water Resour. Res., 49, 3566–3575, doi:10.1002/wrcr.20187, 2013.

Desilets, D., Zreda, M., and Ferré, T.: Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays, Water Resour. Res., 46, W11505, doi:10.1029/2009WR008726, 2010.

Franz, T., Zreda, M., Ferre, T., Rosolem, R., Zweck, C., Stillman, S., Zeng, X., and Shuttleworth, W.: Measurement depth of the cosmic ray soil moisture probe affected by hydrogen from various sources, Water Resour. Res., 48, W08515, doi:10.1029/2012WR011871, 2012a.

Franz, T., Zreda, M., Rosolem, R., and Ferre, T.: Field validation of a cosmic-ray neutron sensor using a distributed sensor network, Vadose Zone J., 11, doi:10.2136/vzj2012.0046, 2012b.

Franz, T., Zreda, M., Rosolem, R., and Ferre, T.: A universal calibration function for determination of soil moisture with cosmic-ray neutrons, Hydrol. Earth Syst. Sc., 17, 453–460, doi:10.5194/hess-17-453-2013, 2013.

Gerrits, A.M.J., Pfister, L., and Savenije, H.H.G.: Spatial and temporal variability of canopy and forest floor interception in a beech forest, Hydrol. Process., 24, 3011–3025, doi:10.1002/hyp.7712, 2010.

Gravano, E., Bussotti, F., Grossoni, P., and Tani, C.: Morpho-anatomical and functional modifications in beech leaves on the top ridge of the Apennines (Central Italy), Phyton Horn, 39, 41-46, 1999.

Gupta, H., Kling, H., Yilmaz, K., and Martinez, G.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80-91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Hawdon, A., McJannet, D., and Wallace, J.: Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia, Water Resour. Res., 50, 5029–5043, doi:10.1002/2013WR015138, 2014.

Hendrick, L.D. and Edge, R.D.: Cosmic-ray neutrons near the Earth, Phys. Rev., 145, 1023-1025, 1966.

Howard, P.J.A., and Howard, D.M.: Use of organic carbon and loss-on-ignition to estimate soil organic matter in different soil types and horizons, Biol. Fertil. Soils, 9, 306-310, 1990.

Iwema, J., Rosolem, R., Baatz, R., Wagener, T., and Bogena, H.R.: Investigating temporal field sampling strategies for site-specific calibration of three soil moisture-neutron intensity parameterisation methods, Hydrol. Earth Syst. Sci., 19, 3203–3216, doi:10.5194/hess-19-3203-2015, 2015.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424-425, doi:10.1016/j.jhydrol.2012.01.011, 2012.

Kodama, M., Kudo, S., and Kosuge, T.: Application of atmospheric neutrons to soil moisture measurement. Soil Sci., 140, 237-242, 1985

Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P., and Zacharias, S.: Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons. Water Resour. Res., 51, 5772-5790, doi:10.1002/2015WR017169, 2015.

Lv, L., Franz, T., Robinson, D., and Jones, S.: Measured and modeled soil moisture compared with cosmic-ray neutron probe estimates in a mixed forest, Vadose Zone J., 13, doi:10.2136/vzj2014.06.0077, 2014.

Ochsner, T., Cosh, M., Cuenca, R., Dorigo, W., Draper, C., Hagimoto, Y., Kerr, Y., Njoku, E., Small, E., and Zreda, M.: State of the art in large-scale soil moisture monitoring, Soil Sci. Soc. Am. J., 77, 1888, doi:10.2136/sssaj2013.03.0093, 2013.

Rivera Villarreyes, C. A., Baroni, G., and Oswald, S. E.: Integral quantification of seasonal soil moisture changes in farmland by cosmic-ray neutrons, Hydrol. Earth Syst. Sci., 15, 3843–3859, doi:10.5194/hess-15-3843-2011, 2011.

Rosolem, R., Shuttleworth, W., Zreda, M., Franz, T., Zeng, X., and Kurc, S.: The effect of atmospheric water vapor on neutron count in the cosmic-ray soil moisture observing system, J. Hydrometeorol., 14, 1659-1671, doi:10.1175/JHM-D-12-0120.1, 2013.

Santa Regina, I., Tarazona, T., and Calvo, R.: Aboveground biomass in a beech forest and a Scots pine plantation in the Sierra de la Demanda area of northern Spain, Ann. Sci. Forest, 54, 261-269, doi:10.1051/forest:19970304, 1997.

Shuttleworth, J., Rosolem, R., Zreda, M., and Franz, T.: The COsmic-ray Soil Moisture Interaction Code (COSMIC) for use in data assimilation, Hydrol. Earth Syst. Sci., 17, 3205–3217, doi:10.5194/hess-17-3205-2013, 2013.

~~Rivera Villarreyes, C. A., Baroni, G., and Oswald, S. E.: Integral quantification of seasonal soil moisture changes in farmland by cosmic-ray neutrons, Hydrol. Earth Syst. Sci., 15, 3843–3859, doi:10.5194/hess-15-3843-2011, 2011.~~

Western, A., Zhou, S.-L., Grayson, R., McMahon, T., Blöschl, G., and Wilson, D.: Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, J. Hydrol., 286, 113-134, doi:10.1016/j.jhydrol.2003.09.014, 2004.

Zreda, M., Desilets, D., Ferré, T., and Scott, R.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, Geophys. Res. Lett., 35, L21402, doi:10.1029/2008GL035655, 2008.

Zreda, M., Shuttleworth, W., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.: COSMOS: the COsmic-ray Soil Moisture Observing System, Hydrol. Earth Syst. Sc., 16, 4079–4099, doi:10.5194/hess-16-4079-2012, 2012.

Table 1. Fractions of different tree stands in percent within the footprint of the CRS.

	Radius 0-50 m	**Radius 50-150 m**	**Radius 150-300 m**	**Total**
Beech	85.2	32.8	48.7	55.5
Pine	3.0	26.3	17.6	15.6
Spruce	5.8	20.9	11.1	12.6
Oak	0.0	10.3	12.5	7.6
Open (grass)	6.0	9.7	3.9	6.5
Larch	0.0	0.0	5.5	1.8
Birch	0.0	0.0	0.7	0.2

Table ~~1~~2. Overview of the f~~ive~~our weighting ~~and correction~~ approaches for other than soil moisture effects on the CRS signal.

Approach	**1 SDW**	**2 DSW**	**3 DDW**	**4 DDWnl**
~~simple depth-weighting~~	~~yes~~	~~no~~	~~no~~	~~no~~
consideration of depth-specific W_L and SOM+B_R ~~separately~~	no	yes	yes	yes
distance depth-weighting	no	no	yes	yes
non-linear depth-weighting	no	no	no	yes
~~consideration of above-ground biomass~~	~~no~~	~~no~~	~~no~~	~~no~~

Table 3~~2~~. Example of depth weighting (DSW) for an ~~critical~~effective measurement depth of z^* = 22.1 cm, a = 0.0903 and b = 1. Calibration campaign date 21 November 2014 (F4). Note the difference in specific weights if only soil water content θ is considered (wt(z,θ)) or if W_L and SOM-+-B_R is also considered (wt(z,H_p)).

Layer (cm)	**θ ($m^3\ m^{-3}$)**	**W_L ($m^3\ m^{-3}$)**	**SOM+B_R ($m^3\ m^{-3}$)**	**H_p ($m^3\ m^{-3}$)**	**ρ_{bd} ($g\ cm^{-3}$)**
0-5	0.187	0.002	0.034	0.223	0.669
5-10	0.136	0.004	0.024	0.163	1.143
10-15	0.117	0.004	0.019	0.140	1.217
15-20	0.109	0.004	0.015	0.129	1.256
20-25	0.106	0.005	0.013	0.124	1.359
25-30	0.100	0.005	0.012	0.118	1.431

z (cm)	$\mathbf{wt(z,\theta)}$	$\int_z^{z+5}\mathbf{wt(z,\theta)}$	$\mathbf{wt(z,}\boldsymbol{H_p}\mathbf{)}$	$\int_z^{z+5}\mathbf{wt(z,}\boldsymbol{H_p}\mathbf{)}$
0	0.079	0.356	0.090	0.401
5	0.063	0.278	0.070	0.299
10	0.048	0.200	0.050	0.197
15	0.032	0.122	0.029	0.095
20	0.017	0.044	0.009	0.009
25	0.001	0.000	0.000	0.000
		Σ=1.00		**Σ=1.00**

Table 4~~3~~. Atmospheric and soil parameters as well as neutron counts for the 10 calibrations. ~~P is the a~~Atmospheric pressure P, ~~p_{v0} is the~~ absolute humidity p_{v0}, ~~N_{raw} is the~~ raw neutron count N_{raw}, ~~N_p is the~~ pressure corrected neutron count N_p, ~~N_{pi} is the~~ pressure and incoming radiation corrected neutron count N_{pi}, ~~N_{pih} is the~~ pressure, incoming radiation and water vapor corrected neutron count N_{pih}, ~~N_0 is the~~ calibration neutron count N_0~~.~~, ~~N_{nm} is the~~ incoming radiation from the neutron monitor N_{nm}, ~~θ_{30cm} is the~~ average soil moisture of the top 30 cm θ_{30cm}, ~~θ_{depthW} is the~~ depth-weighted soil moisture θ_{depthW}, ~~$(W_L+SOM+B_R)_{depthW}$ is the~~ depth-weighted sum of volumetric lattice water content, soil organic matter and root biomass water equivalent $(W_L+SOM+B_R)_{depthW}$, ~~$(H_p)_{depthW}$ is the~~ depth-weighted water equivalent~~hydrogen content~~ of belowground hydrogen pools $(H_p)_{depthW}$, ~~$(\rho_{bd})_{depthW}$ is the~~ depth-weighted bulk density $(\rho_{bd})_{depthW}$ and ~~θ_{mod} is the~~ average volumetric soil water content θ_{mod} of the resulting time series using the N_0-calibration function ~~(Desilets et al., 2010)~~ with standard parameters. Mean (μ) and standard deviation (σ) values of the 10 calibration campaigns are given in the two bottom lines.

Calibration	**P (hPa)**	**p_{v0} (g m^{-3})**	**N_{raw} (~~n h^{-1}~~counts h^{-1})**	**N_p (~~n h^{-1}~~counts h^{-1})**	**N_{pi} (~~n h^{-1}~~counts h^{-1})**	**N_{pih} (~~n h^{-1}~~counts h^{-1})**	**N_0 (~~n h^{-1}~~counts h^{-1})**
Winter	984.0	5.7	606.2	514.9	518.8	509.4	872.4
Spring1	999.3	8.6	549.2	523.0	527.5	526.2	868.7
Spring2	1021.0	4.9	491.1	550.6	542.8	530.5	871.1
Spring3	1002.9	9.6	544.7	533.1	539.9	541.5	869.2
Spring4	1019.0	8.0	503.4	556.0	549.4	546.1	879.0
Summer	1008.7	14.0	613.3	626.6	623.8	640.5	858.2
Fall1	998.7	11.5	624.7	592.4	593.8	601.5	909.5
Fall2	1014.1	7.8	509.3	542.1	546.7	542.8	876.2
Fall3	990.3	8.5	630.4	561.4	580.4	578.5	892.8
Fall4	1016.7	6.6	544.4	591.0	577.7	569.9	885.7
μ	**1005.5**	**8.5**	**561.7**	**559.1**	**560.1**	**558.7**	**878.3**
σ	**11.9**	**2.6**	**50.2**	**33.1**	**31.1**	**37.5**	**13.8**

Kommentar [A113]: Rev3: Insert line!

Kommentar [A114]: Reply: I would love to. Actually there was a line when we submitted the manuscript. But it seems that HESS does not allow lines within tables. However, I will try to convince them to leave this line in there because I also think it is necessary.

Calibration	**N_{nm} (~~n h^{-1}~~counts h^{-1})**	**θ_{30cm} (m^3 m^{-3})**	**θ_{depthW} (m^3 m^{-3})**	**$(W_L+SOM+B_R)_{depthW}$ (m^3 m^{-3})**	**$(H_p)_{depthW}$ (m^3 m^{-3})**	**$(\rho_{bd})_{depthW}$ (g cm^{-3})**	**θ_{mod} (m^3 m^{-3})**
Winter	325.8	0.163	0.228	0.0343	0.262	0.985	0.141

Spring1	325.5	0.153	0.200	0.0340	0.234	1.013	0.143
Spring2	333.0	0.150	0.185	0.0311	0.216	0.955	0.137
Spring3	324.1	0.140	0.175	0.0324	0.207	1.000	0.143
Spring4	332.2	0.139	0.170	0.0302	0.200	0.957	0.145
Summer	329.8	0.073	0.080	0.0278	0.108	1.074	0.151
Fall1	327.4	0.112	0.137	0.0299	0.167	1.016	0.182
Fall2	325.5	0.140	0.174	0.0310	0.205	0.970	0.144
Fall3	317.5	0.119	0.149	0.0316	0.181	1.018	0.166
Fall4	335.8	0.126	0.150	0.0293	0.179	0.981	0.155
μ	**327.7**	**0.131**	**0.165**	**0.0312**	**0.196**	**0.997**	**0.151**
σ	**5.0**	**0.024**	**0.038**	**0.0019**	**0.039**	**0.034**	**0.013**

Table 5~~4~~. Means (*μ*) and standard deviations (*σ*) of calibration parameter N_0 and means (*μ*) and standard deviations (*σ*) of resulting time series of volumetric soil water content θ_{mod} for the four~~ive~~ weighting approaches with 10 calibration campaigns each.

Approach	**$(N_0)_\mu$ (~~n h^{-1}~~counts h^{-1})**	**$(N_0)_\sigma$ (~~n h^{-1}~~counts h^{-1})**	**$(\theta_{mod})_\mu$ (m^3 m^{-3})**	**$(\theta_{mod})_\sigma$ (m^3 m^{-3})**
1 SDW	855.0	17.3	0.158	0.015
2 DSW	878.3	13.8	0.151	0.013
3 DDW	841.9~~0~~	13~~21~~.7~~5~~	0.139~~8~~	0.012~~7~~
4 DDWnl	828~~7~~.1~~7~~	13~~9~~.3~~4~~	0.134~~3~~	0.012~~6~~
~~5 ABC~~	~~1970.9~~	~~50.4~~	~~0.138~~	~~0.017~~

Kommentar [A115]: Rev3: I am not sure how to read these results. Can you clarify in the figure caption or in text (which also glosses over this in one short paragraph at the end of section 4.3).

Kommentar [A116]: Reply: We modified the text to make the table clearer: 'All resulted in largely deviating N0-values between the individual calibrations (see means and standard deviations in column 1 and 2 of Table 5). This in turn led to differences in the time series of volumetric soil water content between the individual calibrations (see means and standard deviations in column 3 and 4 of Table 5).'

Table 6~~5~~. Modified~~New~~ calibration parameters for the four~~ive~~ weighting approaches.

	N_0	**a_0**	**a_1**	**a_2**
1 SDW	926.3	0.203	0.109	0.238
2 DSW	1007.8	0.203	0.114	0.267
3 DDW	810~~4~~.7	0.326~~8~~	0.001	0.310~~1~~
4 DDWnl	779~~04~~.3	0.314~~272~~	0.001~~0~~	0.285~~3~~
~~5 ABC~~	~~1249.1~~	~~0.502~~	~~0.001~~	~~0.312~~

Table ~~6~~7. Performance measures for the four~~ive~~ weighting approaches – comparison of modified~~new~~ calibration (mdf) with standard calibration (stan~~SD~~). KGE’ is the modified Kling-Gupta efficiency, β is the bias ratio and γ is the variability ratio. ~~μ (~~KGE’)$_\mu$ and ~~σ (~~KGE')$_\sigma$ represent the mean and standard deviation of the KGE’ values of the 10 individual single-point standard calibrations.

	KGE' mdf~~new~~	**β mdf~~new~~**	**γ mdf~~new~~**	**~~μ (~~KGE' stan~~SD~~)$_\mu$**	**~~σ (~~KGE' stan~~SD~~)$_\sigma$**	**~~μ (~~β stan~~SD~~)$_\mu$**	**~~μ (~~γ stan~~SD~~)$_\mu$**
1 SDW	0.830	0.849	0.986	0.675	0.045	1.120	1.258
2 DSW	0.880	0.915	0.964	0.727	0.035	1.032	1.231
3 DDW	0.891~~921~~	1.076~~18~~	0~~1~~.986~~006~~	0.712~~676~~	0.081~~7~~	0.878~~7~~	1.237~~58~~
4 DDWnl	0.833~~71~~	1.148~~090~~	1.01~~5~~1	0.681~~74~~	0.096~~107~~	0.8~~2~~18	1.244~~6~~
~~5 ABC~~	~~0.920~~	~~1.025~~	~~0.999~~	~~0.676~~	~~0.087~~	~~0.887~~	~~1.258~~

Table 8. Hydrogen pools (in kg hydrogen per m^2) in the CRS footprint for different moisture conditions (wet: 0.29 m^3 m^{-3}, full canopy and litter storage; intermediate: 0.17 m^3 m^{-3}, dry canopy and moist litter storage; dry: 0.05 m^3 m^{-3}). Above-ground biomass is split into a static part (AGB wet static) comprising stem, branches and dry litter and a variable part (AGB wet variable) that represents leaves.

Hydrogen Pool	**Wet (kg m^{-2})**	**Intermediate (kg m^{-2})**	**Dry (kg m^{-2})**
AGB wet static	5.24	5.24	5.24
AGB wet variable	0.22	0.22	0.22
SOM+R_B	0.36	0.44	0.66
Lattice water	0.05	0.07	0.15
Pore water	4.12	3.26	1.77
Litter water	0.31	0.11	0.00
Interception	0.17	0.00	0.00
Total	10.47	9.35	8.04

Figure 1. ~~Field site location in Müritz National Park in north-eastern Germany.~~Soil sampling locations for calibration (blue dots) and forest vegetation around the CRS (red dot in the center). The TDT soil moisture sensors are located in close vicinity to the sampling locations. The larger yellow circle approximates the footprint of the CRS as it was assumed when sampling took place (diameter approximately 300 m). The smaller yellow circle approximates the footprint of the CRS according to newer modeling results by Köhli et al. (2015) (diameter approximately 200 m). Inset: Field site location in Müritz National Park in north-eastern Germany.

~~Figure 2. Soil sampling locations for calibration (blue dots) and forest vegetation around the CRS (red dot in the center). The TDT soil moisture sensors are located in close vicinity to the sampling locations. The yellow circle approximates the footprint of the CRS (diameter = 300 m).~~

Figure ~~4~~2. Simplified representation of f~~F~~actors influencing the raw neutron count (N_{raw}) and the measurement support of the CRS in terms of ~~critical~~effective measurement depth and footprint. Temporally variable factors are shown on the left: b~~B~~arometric pressure (P), canopy interception (I), air humidity (H) and litter layer interception (L). Temporally constant factors (for our study site) are shown on the right: vegetation above and below the sensor (V), soil organic matter (SOM), root biomass (B_R) and lattice water (W_L)~~., vegetation (V), litter layer (L), soil organic matter (SOM), root biomass (B_R) and lattice water (W_L)~~ All these factors need to be accounted for in order to isolate the~~ signal from~~ soil water content signal (θ). The time-variable factors require permanent monitoring and dynamic correction, the influence of the constant factors is taken into account during calibration. The combination of the time-variable and time-constant factors leads to a site specific temporally variable effective measurement depth and footprint diameter.

~~Figure 5. Mass of hydrogen in individual beech trees in stem and branches (red diamonds) and leaves (green triangles) in relation to diameter at breast height (DBH). Fraction of leaf hydrogen mass of total aboveground tree hydrogen mass (orange line).~~

Kommentar [A117]: Rev1: This map should be integrated in figure 2.

Kommentar [A118]: Done!

Kommentar [A119]: Rev1: According to recent results of Köhli et al. (2015) the footprint is considerably smaller than 300 m. Please adapt the figure. In addition, it would be helpful to color the aerial photograph according to the different tree species.

Kommentar [A120]: Reply: The figure was changed accordingly. Since this figure is also supposed to illustrate our sampling scheme we would like to keep the 'old' footprint size. To make the difference clear we rephrased: 'The yellow circle approximates the footprint of the CRS as it was assumed when sampling took place'. The distribution of different tree species can be seen on Fig. 3 and it would probably make this Figure too busy adding colors or patterns on top of the photograph.

Kommentar [A121]: Rev1: This map should be integrated in figure 2.

Kommentar [A122]: Done!

Kommentar [A123]: Rev1: According to recent results of Köhli et al. (2015) the footprint is considerably smaller than 300 m. Please adapt the figure. In addition, it would be helpful to color the aerial photograph according to the different tree species.

Kommentar [A124]: Reply: Since this figure is also supposed to illustrate our sampling scheme we would like to keep the 'old' footprint size. To make that clear we rephrased: 'The yellow circle approximates the footprint of the CRS as it was assumed when sampling took place'. The distribution of different tree species can be seen on Fig. 3 and it would probably make this Figure too busy adding colors or patterns on top of the photograph.

Figure ~~6~~3. Gravimetrically determined volumetric soil water content patterns in the footprint of the CRS for the 10 calibration dates. The colored dots indicate the unweighted average value from 0 to 30 cm at the 18 calibration locations. Background colors represent the unweighted average value of all 108 soil samples. Different forest stands (pine, beech, oak, spruce) are indicated by the patterned background.

Figure ~~7~~4. Upper panel: volumetric water content derived from CRS data for each of the 10 calibration dates separately (vertical lines indicate calibration dates, colors correspond to time series colors). Filled circles represent the ~~depth-~~weighted volumetric water content at the time of calibration (according to D~~S~~DW). Lower panel: differences in water content between calibration S1 ~~the calibration resulting in the driest time series (S2)~~ and all other calibrations~~.~~ expressed as a percentage of the total possible range of soil water content – ranging from 0.04 m^3 m^{-3} to 0.34 m^3 m^{-3}at our field site~~residual water content to saturated water content~~ (color coding corresponds to calibration dates in the upper panel).

Figure ~~8~~5. Modified~~New~~ calibration functions (solid lines) for the four different weighting approaches (simple depth-weighting SDW, depth-specific weighting DSW, distance-depth-weighting DDW, distance-depth-weighting, non-linear DDWnl), each one derived from 10 calibration points (circles). Calibration points are better captured by flatter calibration functions (solid lines) with modified~~new~~ calibration parameters than by any of the standard calibration functions (dotted lines) based on a single calibration data set only (days S2 and F1 as an example). Black lines illustrate that differences in soil moisture between the results of individual calibrations are larger when soil moisture is high. The inset magnifies the area around the calibration points.

Figure 9~~6~~. Time series of volumetric water content derived with modified~~new~~ calibration functions using~~with new calibration~~ parameters based on the four~~ive~~ calibration approaches: simple depth-weighting (SDW), depth-specific weighting (DSW), distance-depth-weighting (DDW) and~~,~~ distance-depth-weighting, non-linear (DDWnl)~~ and aboveground biomass correction (ABC)~~. Filled circles represent the weighted average of volumetric water content obtained from soil cores at the time of calibration (weighting according to DDW).

Figure 7~~10~~. Average v~~V~~olumetric water content derived from TDT measurements (black line) and CRS measurements (orange line) using different calibration functions. Upper panel: the orange line is an average of the volumetric water content derived from the 10 calibration campaigns of the CRS using the standard N_0-calibration function from Desilets et al. (2010) applying the DDW ~~weighting~~ approach. Grey dotted lines are results for 10 individual calibration campaigns (KGE' values range from 0.579 to 0.834). Lower panel: the orange line is the volumetric water content derived from the~~a~~ ~~new~~ calibration function with modified calibration parameters applying the DDW weighting approach based on all 10 calibration dates. The colored vertical lines mark the days of the last five calibration campaigns.

Figure 8~~11~~. Best-fit N_0-calibration functions (red-brown colored lines) for all combinations of two-point calibrations (blue dots). Best-fit N_0-calibration function for 10-point calibration (black line). Best-fit two-point N_0-calibration function derived from calibration points with highest and lowest volumetric water content (yellow line).

Figure 9~~12~~. Performance of CRS soil water content data derived from two-point calibrations in relation to difference~~stance~~ between soil moisture states ($\Delta\theta$) at the two calibration dates. The color bar indicates volumetric soil water content. Left panel: points are colored according to the soil water content of the drier calibration date. Right panel: points are colored according to the soil water content of the wetter calibration date. Dashed lines indicate~~: Pareto front indicating~~ that soil moisture differences of less than 0.1 m^3 m^{-3} can produce N_0-calibration curves with sub-optimal conversions of neutron counts to volumetric soil water content.

Figure 10. Mass of hydrogen in individual beech trees in stem and branches (red diamonds) and leaves (green triangles) in relation to diameter at breast height (DBH). Fraction of leaf hydrogen mass to~~of~~ total aboveground tree hydrogen mass (orange line).

Figure 11: Varying hydrogen pools in the beech forest surrounding the CRS for three different site conditions. AGB (above-ground biomass) wet variable represents hydrogen contained in deciduous leaves (both in the biomass and in the leaf water). AGB wet static comprises hydrogen contained in biomass and water of tree stems and branches as well as in biomass of the litter layer.

Figure S1. Incoming neutron flux from the neutron monitors in Kiel, Germany and Jungfraujoch, Switzerland and synthetic continuous time series of incoming neutron flux combined from these two and used for the corrections in this study.

Figure S2. Comparison of depth-(and distance-) weighted averages of gravimetrically determined soil water content with unweighted gravimetrically determined soil water content of the upper 15 cm of the soil. The first two weighting approaches overestimate soil water content in the upper 15 cm especially at high soil water contents. The last two approaches have only a slight negative offset and no significant relationship with wetness conditions.

Kommentar [A125]: Rev1: The Pareto front needs to be discussed in the text as well.

Kommentar [A126]: Reply: We realized that it is actually not a Pareto front. So we added to the text: 'The existence of a rather clear front in Fig. 9 indicates that the calibrated neutron count-soil water content conversion will always perform well if the soil moisture differences between the two calibrations are sufficiently large.'

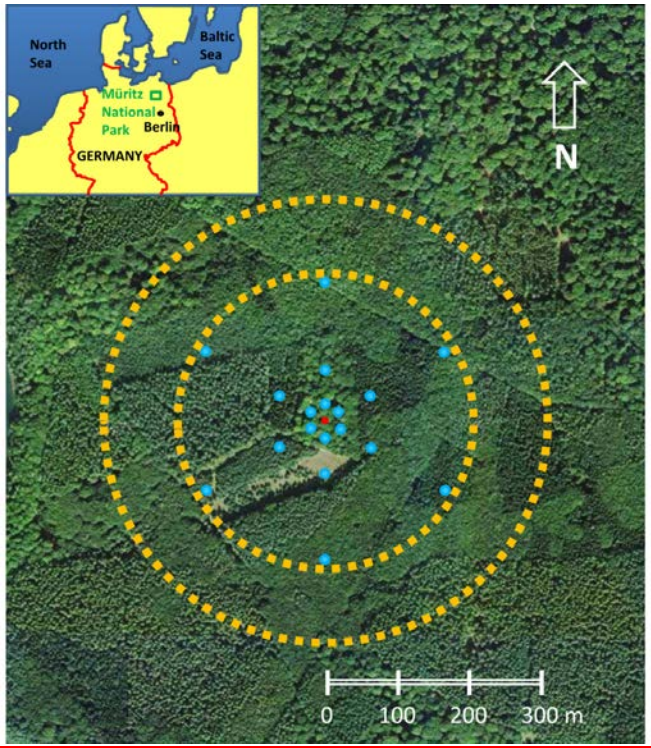

Figure 1. Soil sampling locations for calibration (blue dots) and forest vegetation around the CRS (red dot in the center). The TDT soil moisture sensors are located in close vicinity to the sampling locations. The larger yellow circle approximates the footprint of the CRS as it was assumed when sampling took place (diameter approximately 300 m). The smaller yellow circle approximates the footprint of the CRS according to newer modeling results by Köhli et al. (2015) (diameter approximately 200 m). Inset: Field site location in Müritz National Park in north-eastern Germany.

Kommentar [A127]: Rev1: According to recent results of Köhli et al. (2015) the footprint is considerably smaller than 300 m. Please adapt the figure. In addition, it would be helpful to color the aerial photograph according to the different tree species.

Kommentar [A128]: Reply: The figure was changed accordingly. Since this figure is also supposed to illustrate our sampling scheme we would like to keep the 'old' footprint size. To make the difference clear we rephrased: 'The yellow circle approximates the footprint of the CRS as it was assumed when sampling took place'. The distribution of different tree species can be seen on Fig. 3 and it would probably make this Figure too busy adding colors or patterns on top of the photograph.

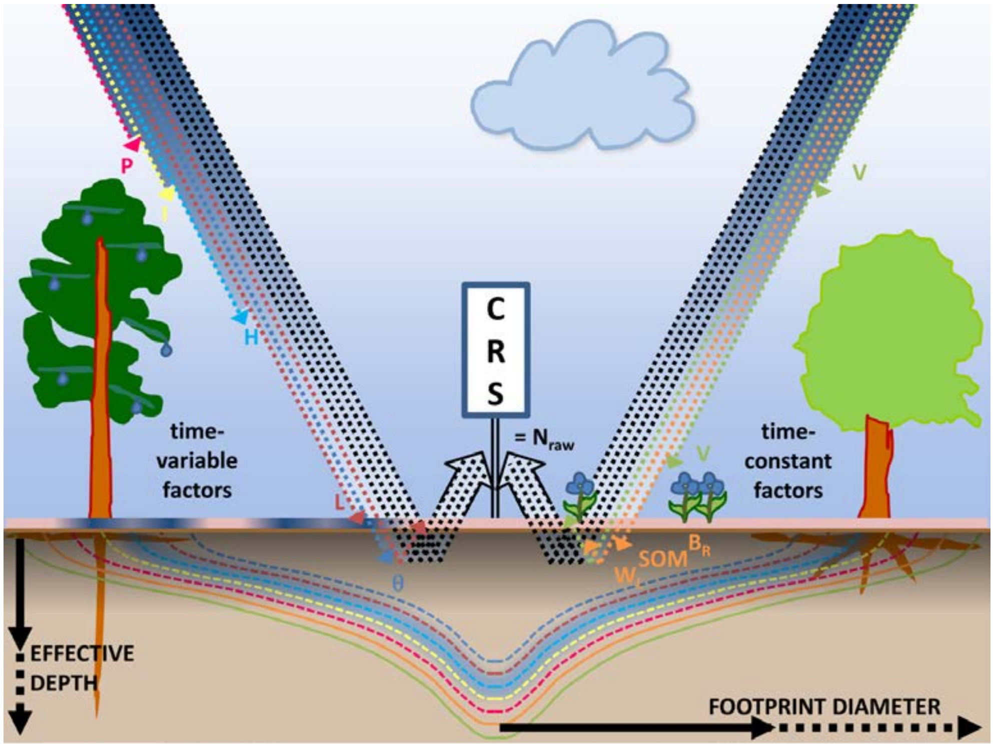

Figure 2. Simplified representation of factors influencing the raw neutron count (N_{raw}) and the measurement support of the CRS in terms of effective measurement depth and footprint. Temporally variable factors are shown on the left: barometric pressure (P), canopy interception (I), air humidity (H) and litter layer interception (L). Temporally constant factors (for our study site) are shown on the right: vegetation above and below the sensor (V), soil organic matter (SOM), root biomass (B_R) and lattice water (W_L). All these factors need to be accounted for in order to isolate the soil water content signal (θ). The time-variable factors require permanent monitoring and dynamic correction, the influence of the constant factors is taken into account during calibration. The combination of the time-variable and time-constant factors leads to a site specific temporally variable effective measurement depth and footprint diameter.

Kommentar [A129]: Rev3: can you group these factors into two groups: (1) constant in time (eg, WL), and (2) variable in time (eg, water vapor)? And then perhaps also into another groups: (3) those measured at calibration (eg, WL), and (4) those measured as time series (eg, water vapor)? The significance is that some parameters are easy to handle, others are difficult, and the two should not be commingled.

Kommentar [A130]: Reply: We revised the figure according to your suggestions. We grouped the factors into two groups and mentioned which of them need to be monitored continuously and which can be accounted for during calibration.

Kommentar [A131]: Rev1: This schematic figure is wrong in presenting the cosmic-ray neutron intensities as actual rays that are reflected by the soil. The actual processes leading to neutron intensity are far more complex (see e.g. Köhli et al., 2015) and should not be presented in this way in a scientific paper. Also the above ground and below ground footprints are not connected in the simple way as suggested by the schematic drawing. Thus, the figure should be omitted.

Kommentar [A132]: Reply: We acknowledge the fact that this figure simplifies the actual processes a lot and have added this statement to the caption. We have modified the figure to resolve two of your concerns. The neutron intensities are not any longer depicted as rays and the above-ground footprint was removed entirely. We still think that the figure helps to get an overview over the many parameters that have to be accounted for before/during the use of the CRS method.

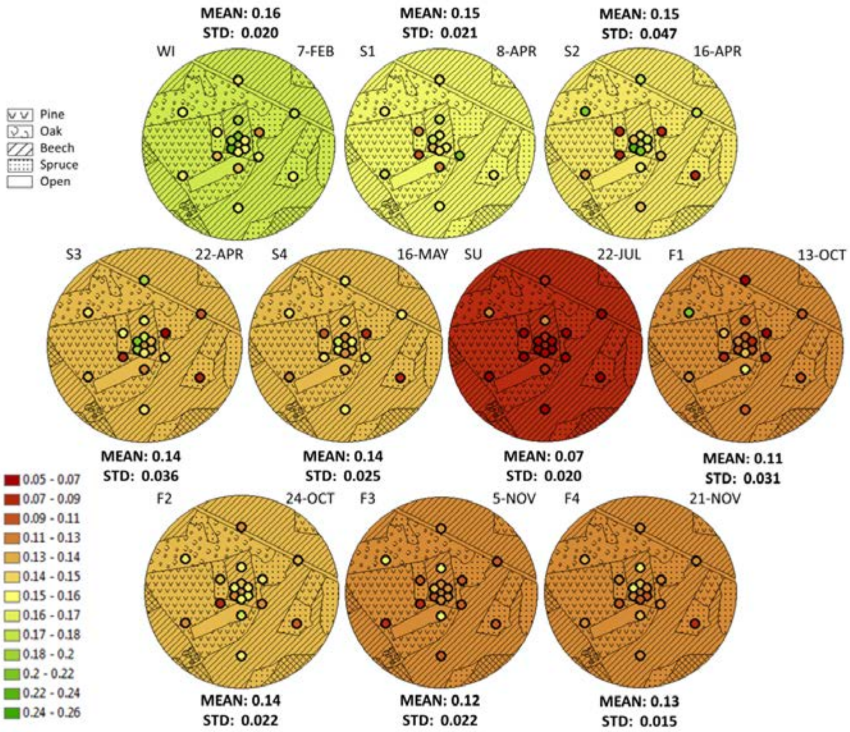

Figure 3. Gravimetrically determined volumetric soil water content patterns in the footprint of the CRS for the 10 calibration dates. The colored dots indicate the unweighted average value from 0 to 30 cm at the 18 calibration locations. Background colors represent the unweighted average value of all 108 soil samples. Different forest stands (pine, beech, oak, spruce) are indicated by the patterned background.

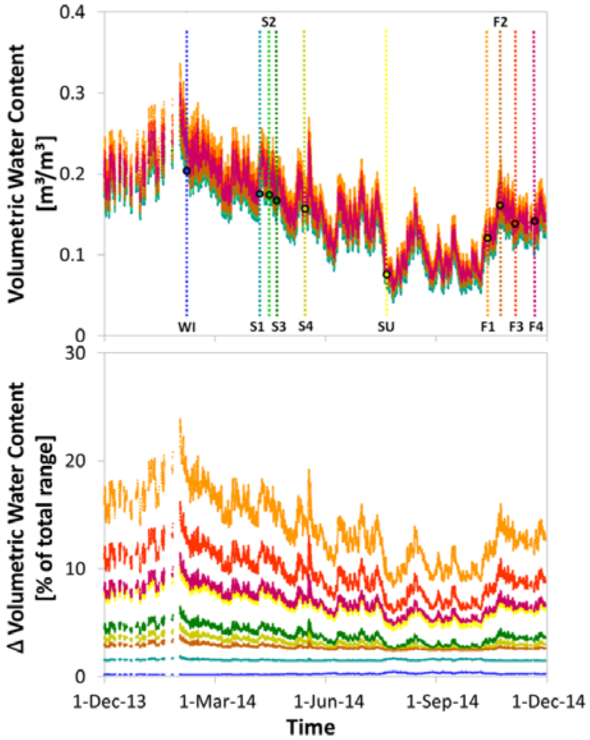

Figure 4. Upper panel: volumetric water content derived from CRS data for each of the 10 calibration dates separately (vertical lines indicate calibration dates, colors correspond to time series colors). Filled circles represent the weighted volumetric water content at the time of calibration (according to DDW). Lower panel: differences in water content between calibration S1 and all other calibrations expressed as a percentage of the total possible range of average soil water content – ranging from 0.04 m^3 m^{-3} to 0.34 m^3 m^{-3}~~residual water content to saturated water content~~at our field site (color coding corresponds to calibration dates in the upper panel).

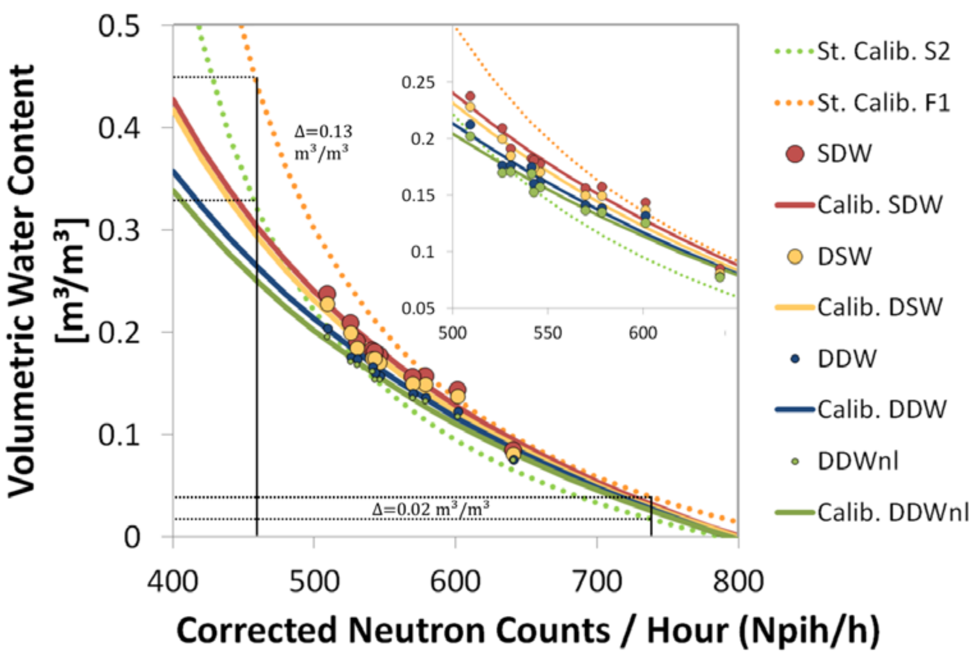

Figure 5. Modified calibration functions (solid lines) for the four different weighting approaches (simple depth-weighting SDW, depth-specific weighting DSW, distance-depth-weighting DDW, distance-depth-weighting, non-linear DDWnl), each one derived from 10 calibration points (circles). Calibration points are better captured by flatter calibration functions (solid lines) with modified calibration parameters than by any of the standard calibration functions (dotted lines) based on a single calibration data set only (days S2 and F1 as an example). Black lines illustrate that differences in soil moisture between the results of individual calibrations are larger when soil moisture is high. The inset magnifies the area around the calibration points.

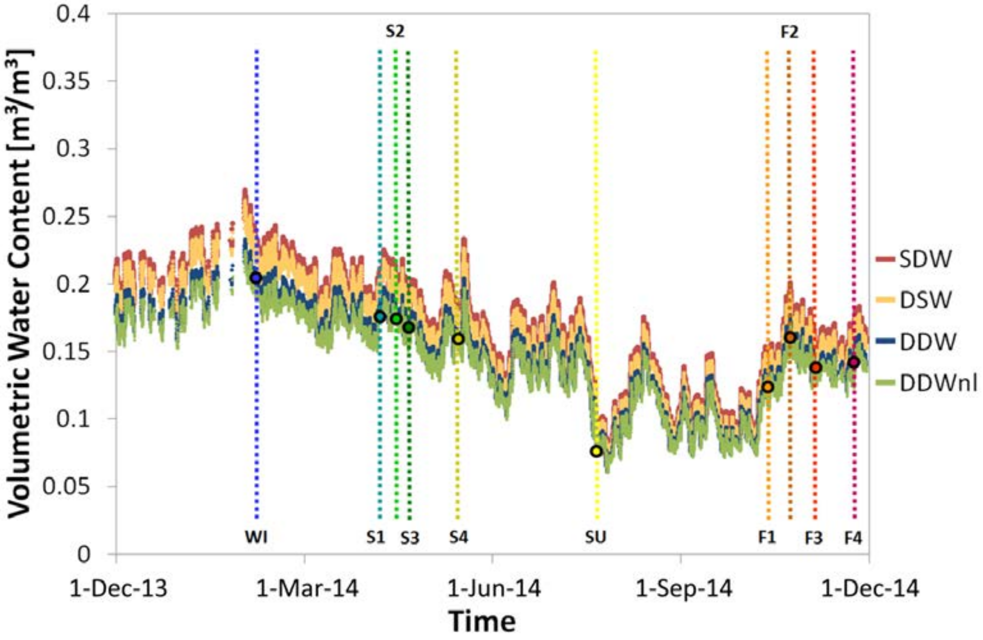

Figure 6. Time series of volumetric water content derived with modified calibration functions using parameters based on the four calibration approaches: simple depth-weighting (SDW), depth-specific weighting (DSW), distance-depth-weighting (DDW) and distance-depth-weighting, non-linear (DDWnl). Filled circles represent the weighted average of volumetric water content obtained from soil cores at the time of calibration (weighting according to DDW).

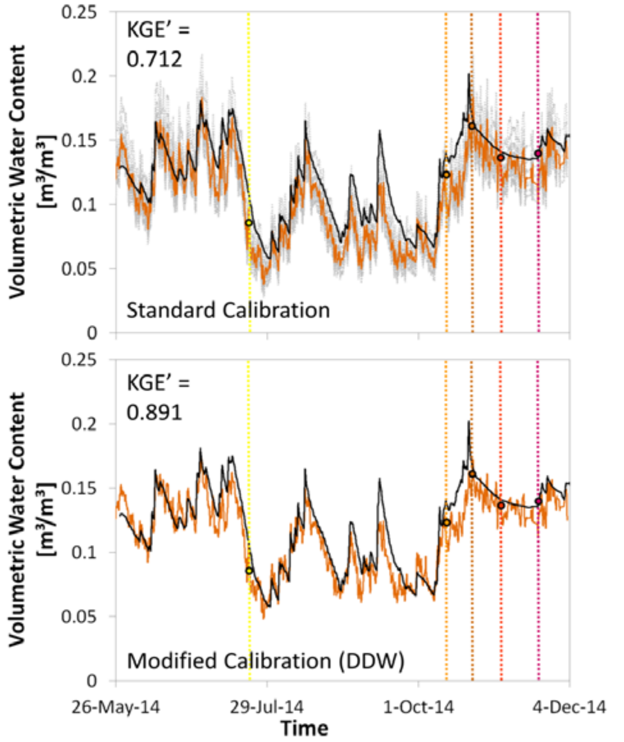

Figure 7. Average volumetric water content derived from TDT point measurements (black line) and CRS measurements (orange line) using different calibration functions. Upper panel: the orange line is an average of the volumetric water content derived from the 10 calibration campaigns of the CRS using the standard N_0-calibration function from Desilets et al. (2010) applying the DDW approach. Grey dotted lines are results for 10 individual calibration campaigns (KGE' values range from 0.579 to 0.834). Lower panel: the orange line is the volumetric water content derived from the calibration function with modified calibration parameters applying the DDW weighting approach based on all 10 calibration dates. The colored vertical lines mark the days of the last five calibration campaigns.

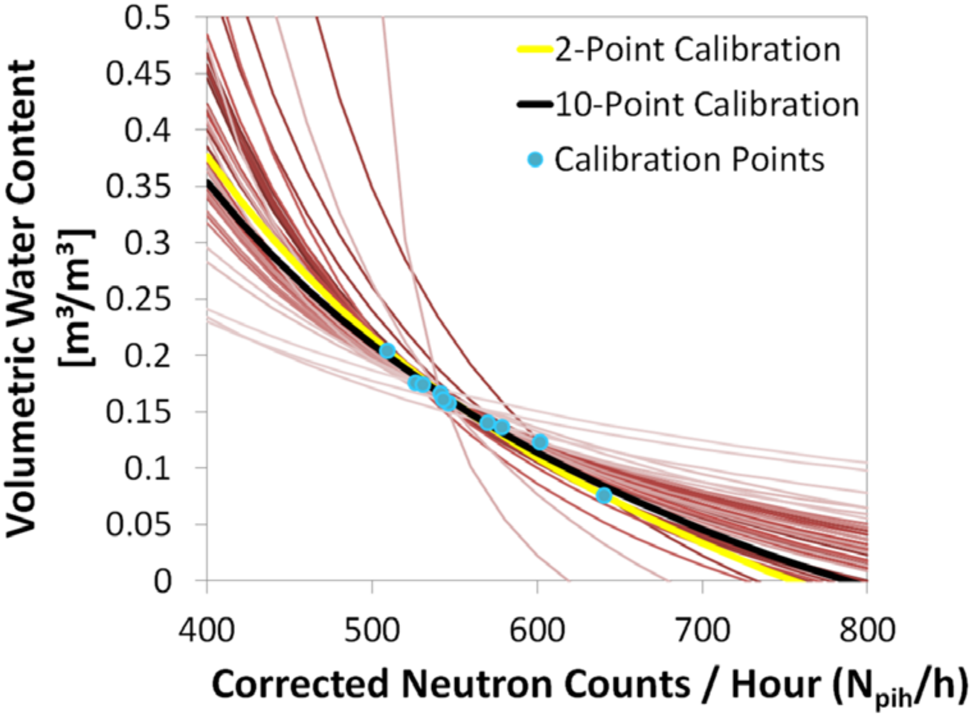

Figure 8. Best-fit N_0-calibration functions (red-brown colored lines) for all combinations of two-point calibrations (blue dots). Best-fit N_0-calibration function for 10-point calibration (black line). Best-fit two-point N_0-calibration function derived from calibration points with highest and lowest volumetric water content (yellow line).

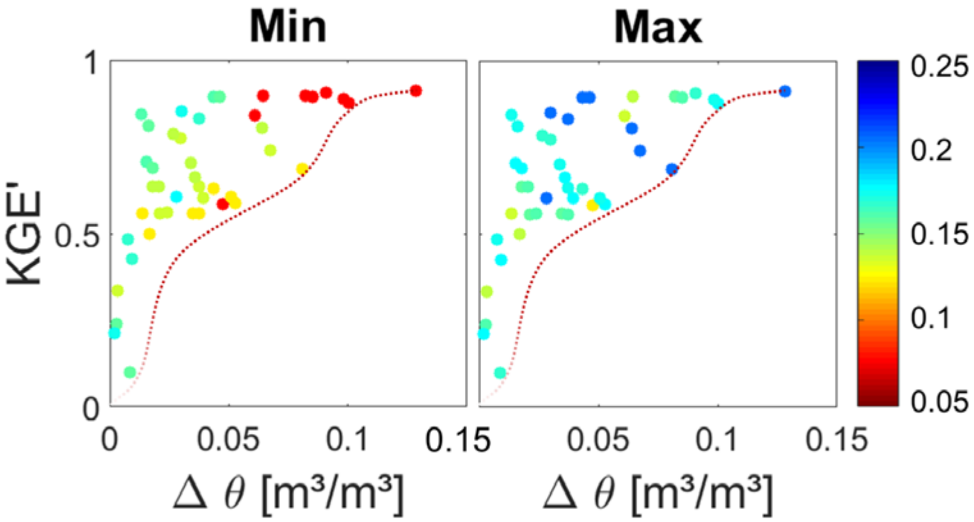

Figure 9. Performance of CRS soil water content data derived from two-point calibrations in relation to difference between soil moisture states ($\Delta\theta$) at the two calibration dates. The color bar indicates volumetric soil water content. Left panel: points are colored according to the soil water content of the drier calibration date. Right panel: points are colored according to the soil water content of the wetter calibration date. Dashed lines indicate that soil moisture differences of less than 0.1 m^3 m^{-3} can produce N_0-calibration curves with sub-optimal conversions of neutron counts to volumetric soil water content.

Kommentar [A133]: Rev1: The Pareto front needs to be discussed in the text as well.

Kommentar [A134]: Reply: We realized that it is actually not a Pareto front. So we added to the text: 'The existence of a rather clear front in Fig. 9 indicates that the calibrated neutron count-soil water content conversion will always perform well if the soil moisture differences between the two calibrations are sufficiently large.'

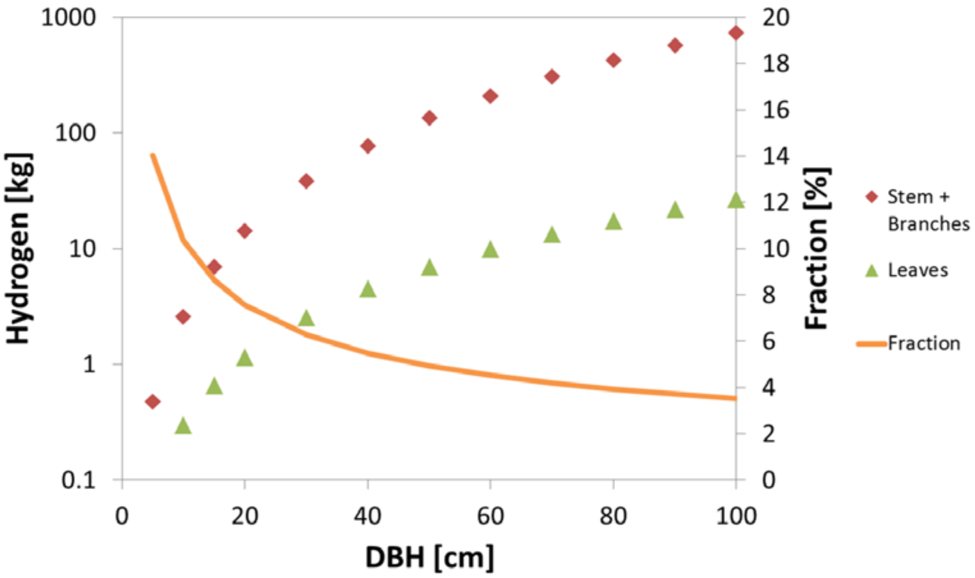

Figure 10. Mass of hydrogen in individual beech trees in stem and branches (red diamonds) and leaves (green triangles) in relation to diameter at breast height (DBH). Fraction of leaf hydrogen mass to~~of~~ total aboveground tree hydrogen mass (orange line).

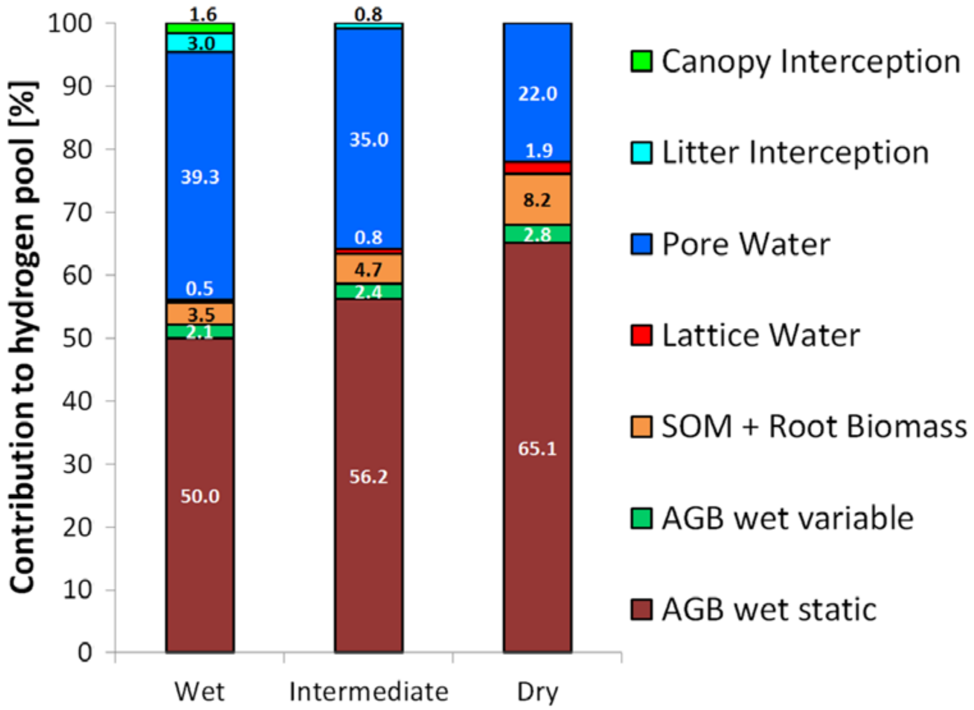

Figure 11: Varying hydrogen pools in the beech forest surrounding the CRS for three different site conditions. AGB (above-ground biomass) wet variable represents hydrogen contained in deciduous leaves (both in the biomass and in the leaf water). AGB wet static comprises hydrogen contained in biomass and water of tree stems and branches as well as in biomass of the litter layer.

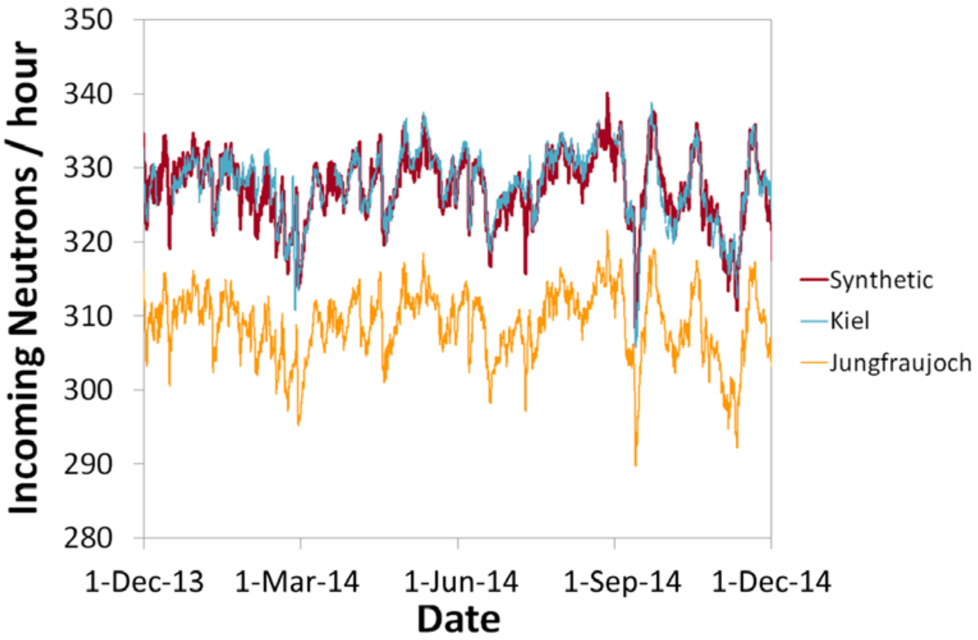

Figure S1. Incoming neutron flux from the neutron monitors in Kiel, Germany and Jungfraujoch, Switzerland and synthetic continuous time series of incoming neutron flux combined from these two and used for the corrections in this study.

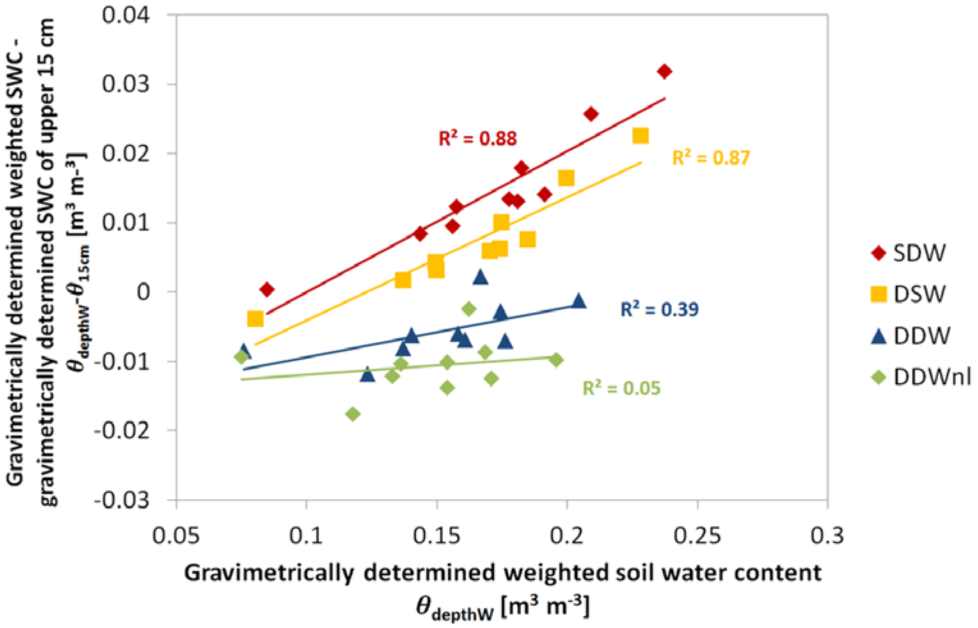

Figure S2. Comparison of depth-(and distance-) weighted averages of gravimetrically determined soil water content with unweighted gravimetrically determined soil water content of the upper 15 cm of the soil. The first two weighting approaches overestimate soil water content in the upper 15 cm especially at high soil water contents. The last two approaches have only a slight negative offset and no significant relationship with wetness conditions.