Response to Interactive comment by H. Bogena (Referee)

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

I have reviewed this manuscript for the second time. The authors have improved their manuscript considering most of my earlier comments. It will be a nice contribution to the HESS. However, there are still several minor issues which need to be considered:
We would like to thank this reviewer for assessing our manuscript for a second time. We were fortunate to have three reviewers that took their job seriously and genuinely cared about the quality of the work and the subject matter.

As mentioned in my earlier review, 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. Also the authors have stated in their cover letter that this procedure was applied in this study but this procedure is still not clearly stated in the revised manuscript and it is also not mentioned in their "Best practice" description.
Thank you for reminding us of this missing piece of information. We added in line 231: 'Calibration has to be performed using the total belowground hydrogen pool (including hydrogen contributions from lattice water (WL), soil organic matter (SOM) and root biomass (BR)). Soil water content is then computed by subtracting the other hydrogen pools from the measured neutron-derived signal.'

In my first review I have stated the following: "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."
The authors answer was: "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)."
Clearly the authors didn't fully grasp my comment. The value of the standard N0-method lies in its predefined form that makes calibration using data from a single sampling date more stable than the modified N0-method. This should be added to the discussion since there might be cases where only one calibration date is available.
Ok, we acknowledge the fact that the standard N0 method does have its value when using only a single calibration. We added in line 587 of the discussion: 'The value of the

standard N0-calibration method becomes apparent when there is only data available from one calibration date. Due to the fixed shape of its calibration function, the general dynamics of the soil water content will still be reproduced.'

L47: …neutrons interact with air and soil nuclei…
Done.

L414: delete "resolution"
Deleted.

Figure 1: The blue dots are difficult to discern.
We turned them into white dots.

Figure 2: The modified schematic shown in this figure is still confusing. Why should the factors influencing the cosmic-ray intensity a ground level be symbolized a rays of different length? I find this irritating and not helpful for the understanding of the involved processes. In addition, the presented sensing depth boundaries are too numerous be discernible and the distribution is not correct. A better representation can be found in Köhli et al., 2015.Thus, in my opinion the figure should be omitted.
We modified the figure to address your concerns. We removed the rays of different length and instead used gradually less intense background color to represent the decreasing number of fast neutrons due to the various factors. We removed all depth boundaries only keeping the maximum extent and one example of reduced depth and footprint. We also modified the shape of the depth boundary so that it resembles better what Köhli et al. (2015) describe.

Response to Interactive comment by Anonymous Referee #2

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

The authors have addressed my comments and concerns.
We would like to thank this reviewer for assessing our manuscript for a second time. We were fortunate to have three reviewers that took their job seriously and genuinely cared about the quality of the work and the subject matter.

Response to Interactive comment by Anonymous Referee #3

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

This paper describes experimental work whose goal was to calibrate cosmic-ray soil moisture probe used in the presence of variable vegetation. I have reviewed this paper before and recommended it to be published with minor revisions. This second submission is much improved. My recommendations, as well as those of other reviewers, have been taken seriously and the improved manuscript is very good. I recommend that it be published, subject to minor technical revisions as described in the attached annotated manuscript.
We would like to thank this reviewer for assessing our manuscript for a second time. We were fortunate to have three reviewers that took their job seriously and genuinely cared about the quality of the work and the subject matter.

L16: give max deviation in absolute values and also as % of the total range
Added: '…deviations of up to 0.1 m^3 m^{-3} (24 % of the total range)…'.

L20: change to "correctly to define"
Changed to: 'A two-point calibration was found to effectively define the shape of the modified calibration function…'.

L48: "fast neutrons" means specific energy (around 1 MeV), so there is no need to qualify this.
Removed.

L603: this is not a calibration method, but a model of neutron intensity that can be used in data assimilation schemes - please clarify this in text.
Changed to: '…another parameterization method developed by Shuttleworth et al. (2013) (the COSMIC operator, a model of neutron intensity used in data assimilation schemes)…'

L660: I am not sure if this is best practice. It is yours, but to be best it should be confirmed by others. I suggest deleting this appendix. Or, at the very least, present it as a proposal for consideration by users.
You are right in that we should not call it 'best practice'. This would indeed be overconfident. Therefore we changed the name to 'Proposed method for calibration…' We would, however, like to keep the Appendix in the manuscript since it provides the sequence of necessary steps that first-time users can follow when they set up and calibrate the instrument. When we started we would have liked to find such a condensed and structured compilation.

Table 1: Totals are not weighted by areas! Area 0-50 m is much smaller than that in 50-150 m, which is smaller than 150-300 m.

You are right. We should have specified this before: ‘The total represents a distance weighted average with an exponential decay towards more distant areas (approximating the exponential distance-weighting from Zreda et al. (2008)).’

Figure 2: support volume may be a better term here
Added ‘volume’.

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 cosmic ray neutrons is a promising technique 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 calibration 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 0.1 m^3 m^{-3} (24 % of the total range) 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 calibration function with a different shape that can vary from one location to another. A two-point calibration was found to effectively ~~proved to be adequate to correctly~~ define the shape of the modified calibration function if the calibration points were taken during both dry and wet conditions spanning at least 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 negligible factor for calibration because the

Kommentar [A1]: Ref3: give max deviation in absolute values and also as % of the total range

Kommentar [A2]: Ref3: change to "correctly to define"

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

1. Introduction

Determining average soil moisture content over larger areas is difficult, mainly for two reasons. Firstly, soil moisture can be highly variable even at small spatial scales, especially under intermediate 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 time-consuming and expensive. 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, 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 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 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 nuclei 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 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.

Kommentar [A3]: Ref1: neutrons interact with air and soil nuclei

Kommentar [A4]: Ref3: "fast neutrons" means specific energy (around 1 MeV), so there is no need to qualify this.

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 with 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 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), 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 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).

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 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 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) 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) explained 87 % of the variation proposing 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 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. 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 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 regression factor that can be derived from calibration or 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 of other variables (such as air pressure and humidity variations or fluctuations in incoming neutron counts). On the other 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 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. 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 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 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. 1

148 and also Fig. 3 for a map of the forest stands and Table 1 for fractions of the different tree stands
149 within the footprint). Depending on the tree species, the mineral soil is covered by an organic soil
150 layer and a litter layer of variable depth and water holding capacity.

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

161

162 ## 3. Methods

163 ### 3.1. Calibration

164 We conducted a total of 10 calibration campaigns throughout one calendar year (2014). The first
165 one (WI) took place in February during winterly conditions with very wet soils. The next four
166 calibrations (S1-4) followed in spring (April-May) and covered the entire period of tree foliation.
167 The sixth calibration (SU) was done under very dry conditions in July and the last four
168 calibrations (F1-4) in fall (October-November) covering the trees' defoliation. For all the
169 calibration campaigns we followed the recommended sampling pattern for the calibration of CRS
170 which was developed by Zreda et al. (2012) and slightly modified and detailed in Franz et al.
171 (2012b). The sampling pattern prescribes 3 concentric circles around the CRS with radii of 25, 75
172 and 200 m, respectively (Fig. 1). The 3 circles are intersected by 6 straight lines that point from
173 the sensor towards north (0°), north-east (60°), south-east (120°), south (180°), south-west (240°)
174 and north-west (300°). Samples are taken in the vicinity of all intersections – the samples do not
175 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 18 samples 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).

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:

$$N_p = N_{raw} * e^{\left(\frac{x-x_0}{L}\right)} \quad (1),$$

with N_p being the pressure corrected neutron counts (counts h^{-1}), N_{raw} the raw neutron counts (counts h^{-1}), x the atmospheric shielding depth (g cm^{-2}) for every time step (derived from atmospheric pressure measured directly inside the CRS case), x_0 the average atmospheric shielding depth (g cm^{-2}) 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):

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

with N_{pi} being the sensor neutron count rate corrected for changes in atmospheric pressure and incoming neutrons (counts h^{-1}), N_{avg} the average count rate of incoming neutrons (counts h^{-1}) over the entire measurement period and N_{nm} the neutron count rate of the neutron monitor for each time step (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 counts h^{-1}) to the mean of the Kiel time series (327 counts h^{-1}). The resulting time series resembles the Kiel time series very closely (Fig. S1).

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

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

with N_{pih} being the sensor neutron count corrected for changes in pressure, incoming neutrons and water vapor (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

230 = 0.115) were determined via neutron transport simulations and a fourth one (N_0) that serves as a
231 calibration parameter accounting for site and sensor specific variations and representing neutron
232 counts over dry soil at reference conditions during calibration. Calibration has to be performed
233 using the total belowground hydrogen pool (including hydrogen contributions from lattice water
234 (W_L), soil organic matter (SOM) and root biomass (B_R)). Soil water content is then computed by
235 subtracting the other hydrogen pools from the measured neutron-derived signal:

236 $$\theta(t) = \left\{\left[a_0 * \left(\frac{N_{pih}(t)}{N_0} - a_1\right)^{-1} - a_2\right] * \rho_{bd}\right\} - W_L - (SOM + B_R) \quad (4).$$

237 The other parameters ρ_{bd}, W_L, SOM and B_R can be measured directly from the calibration soil
238 samples: the bulk density of the soil (ρ_{bd} in g cm^{-3}), the summed volume fraction of lattice water
239 in the soil grains and tightly bound water (W_L in m^3 m^{-3}), the combined volume fraction of soil
240 organic matter and root biomass water equivalent ($SOM+B_R$ in m^3 m^{-3}). In order to calibrate the
241 sensor one first has to determine the depth- (and distance-) weighted averages for ρ_{bd}, W_L,
242 $SOM+B_R$ and θ as well as N_{pih} (averaged over 12 h) for the time of calibration. This is necessary
243 because several factors can influence the effective measurement depth z^* (which is the depth of
244 the soil layer up to which 86 % of the neutrons that the CRS detects originate from) and the
245 footprint size of the sensor (Fig. 2). Afterwards N_0 is adjusted iteratively (e.g. with a simple
246 Solver routine in Microsoft Excel) until the right-hand side of the equation equals the left-hand
247 side.

248 We tested four soil moisture weighting approaches (Table 2), described in detail below, to
249 determine which information is necessary for an accurate calibration.

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

253 $$\begin{cases} wt(z) = a\left[1 - \left(\frac{z}{z^*}\right)^b\right] & 0 \le z \le z^* \\ wt(z) = 0 & z > z^* \end{cases} \quad (5),$$

254 where

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

and

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

and

$$H_p = W_L + SOM + B_R + \theta \quad (8).$$

In these equations z is the soil depth below the surface in cm and z^* is the effective measurement depth in cm, a is a parameter that ensures that the weights are conserved, b controls the curvature of the weighting function and equals 1 for linear weighting, H_p is the water equivalent of the belowground hydrogen pools ($m^3\ m^{-3}$), W_L is lattice water ($m^3\ m^{-3}$), SOM is soil organic matter water equivalent ($m^3\ m^{-3}$), B_R is root biomass water equivalent ($m^3\ m^{-3}$) 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 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 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 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 effective measurement depth decreases with distance from the sensor. The effective measurement depth z^* is calculated according to:

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

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}) (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):

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

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, 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):

$$B_S = 0.0894 * DBH^{2.4679} \quad (11),$$

$$B_B = 0.0317 * DBH^{2.3931} \quad (12),$$

$$B_L = 0.0145 * DBH^{1.9531} \quad (13).$$

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). We first calculated the water mass (W_{agb}) in stems, branches and leaves (assuming a leaf water content of 0.6 kg per kg 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 H_2O) and in dry biomass (M_b = 0.0622 kg H per kg Cellulose: $C_6H_{10}O_5$) the total hydrogen mass (H_{agb}) of above-ground biomass in the beech stand was derived:

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

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.

3.3. Validation

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):

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

With correlation coefficient *r*:

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

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. 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) 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 ($wSOM+B_R$) 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.

4.3. 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 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 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}, and then Eq. (4) to identify N_0 for each calibration. Table 3 provides an example of the depth-weighting following approach 2 (DSW with depth-specific values of W_L and $SOM+B_R$).

381 The values in Table 3 result in a depth-weighted average volumetric water content θ_{depthW} of
382 0.150 m^3 m^{-3}, a depth-weighted water equivalent of belowground hydrogen pools $(H_p)_{depthW}$ of
383 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 $SOM+B_R$
384 were not considered, the values for θ_{depthW} and $(\rho_{bd})_{depthW}$ would change to 0.146 m^3 m^{-3} and
385 1.013 g cm^{-3} respectively, because the effective measurement depth z^* increases when the higher
386 amounts of $SOM+B_R$ in the shallow layers are not considered, thus giving more weight to low
387 soil moisture values in deeper soil horizons.

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

390 Following the standard N_0-calibration approach of Desilets et al. (2010), we should have ended
391 up with the same N_0 value for each of the 10 calibrations. However, the N_0 range we found was
392 considerable – e.g. from 808 to 895 counts h^{-1} for the DDW approach (mean: 841.9 counts h^{-1},
393 standard deviation: 13.7 counts h^{-1}). As a consequence, the 10 computed time series also showed
394 differences in volumetric soil water content (Fig. 4 illustrates results for the DDW approach). In
395 the most extreme case, these differences were larger than 0.1 m^3 m^{-3} (which is equal to 24 % of
396 the total range of soil water content at the site).

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

402 ### 4.4. Modified calibration function

403 To include all information of our 10 calibration campaigns into our analysis, we fitted modified
404 calibration functions to four sets of 10 calibration points derived from the four different
405 weighting approaches (see section 3.1). This was done by using the Microsoft Excel Solver
406 software to optimize the three shape parameters (a_0, a_1, a_2) and N_0 through the calibration point
407 cloud (solid lines in Fig. 5). Plotting the N_{pih}-values of all 10 calibrations against the
408 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. 5) our calibration points are better captured by less steep functions (solid lines in 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 than one would expect.

Kommentar [A5]: Ref1: delete "resolution"

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

4.5. Validation

We tested whether the modified 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 ($\theta_{15\mathrm{cm}}$) (Fig. S2). We found strong linear correlations for two of the weighting approaches (SDW and DSW) with CRS water content being larger than the $\theta_{15\mathrm{cm}}$ 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 approaches 3 and 4 (DDW and DDWnl) offsets of 0.006 and 0.011 m^3 m^{-3} indicated slightly lower weighted 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 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), 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). The linear depth weighting worked better than the non-linear depth weighting (KGE'(DDWnl) = 0.83). 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 to 0.83 with a mean KGE' of 0.71 ($\pm$0.08). It is important to note that all of the modified 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. 7). This supports the hypothesis that at our field site larger than expected changes in neutron count are already caused by subtle changes in soil moisture.

4.6. Optimizing calibration efforts

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. 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 0.1 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. 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.

4.7. Variability of hydrogen pools

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 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 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.

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 soil water content of about 0.031 m^3 m^{-3} (under wet conditions) and 0.018 m^3 m^{-3} (under dry conditions). These differences for wet and dry conditions are due to the fact that the 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 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 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 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).

5.2. Weighting approaches

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 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 the use of a linear depth-weighting function.

5.3. Calibration function

The differences in calibration results are 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 were 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 samplings. 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. The value of the standard N_0-calibration method becomes apparent when there is only data available from one calibration date. Due to the fixed shape of its calibration function, the general dynamics of the soil water content will still be reproduced.

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 we are dealing with the lowest number of counts of all published studies (average N_0 = 878 counts h^{-1}, Table 4). Although the calibration function was theoretically developed for all environments it has not 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. 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 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. Baatz et al. (2014) working also in a low-count environment (N_0: 936 to 1242 counts h^{-1}) with land use ranging from grassland to agriculture to forest compared the standard N_0-calibration method to another parameterization~~calibration~~ method developed by Shuttleworth et al. (2013) (the COSMIC operator, a model of neutron intensity used in data assimilation schemes) and found that the former interpreted dry periods drier and wet periods wetter – which is in

Kommentar [A6]: Ref3: this is not a calibration method, but a model of neutron intensity that can be used in data assimilation schemes - please clarify this in text.

615 accordance to our findings that suggest that the standard N_0-calibration function is too steep. Lv
616 et al. (2014), in a study at a mixed-forest/grassland site also recommended more than one
617 calibration. They operated in a high-count environment in Utah (N_0 = 2189 counts h^{-1}) and
618 attributed the different shape of their calibration function to binary soil moisture patterns at their
619 site where the grassland soils were much drier than the forest soils under wet conditions but just
620 as dry under dry conditions. Our field site is subject to similar spatial variability since it is also
621 comprised of multiple areas with non-uniform soil water content (mean values of soil water
622 contents differ between different forest stands). Following the argumentation of Lv et al. (2014),
623 the fact that distance weighting improved our results can be regarded as an indication that non-
624 homogeneous soil moisture conditions indeed lead to changes in the shape of the calibration
625 function. At our site, distance weighting reduced the spatial variability within the footprint of the
626 sensor since it assigned higher weights to the closest sampling sites which were all located in the
627 homogenous and relatively wet beech forest, while the influence of the drier soils under the
628 coniferous trees was reduced.

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

639

640 ## 6. Conclusion

641 Our results suggest that a one-time calibration of the CRS using the available neutron count
642 corrections and weighting approaches is not sufficient at our field site. This is mainly due to the
643 fact that the shape of the standard N_0-calibration function is not able to reproduce 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 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 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 seasonal changes in above-ground biomass in the forest were found to be negligible. While there is no doubt that further investigations 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, 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.

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 the sensor (Köhli et al, 2015). With the footprint shrinking and expanding and the effective measurement depth in 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.

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

Kommentar [A7]: Ref3: I am not sure if this is best practice. It is yours, but to be best it should be confirmed by others. I suggest deleting this appendix. Or, at the very least, present it as a proposal for consideration by users.

Kommentar [A8]: Reply: You are right in that we should not call it 'best practice'. This would indeed be overconfident. Therefore we changed the name to 'Proposed practice for calibration…' We would, however, like to keep the Appendix in the manuscript since it provides the sequence of necessary steps that first-time users can follow when they set up and calibrate the instrument. When we started we would have liked to find such a condensed and structured compilation.

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).
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. 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 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. 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. Determine the water equivalent of the average hydrogen content of belowground hydrogen pools (H_p) for each soil depth.
    Equation (8).
12. Apply a linear weighting function to your gravimetrically determined H_p measurements accounting for the change in the 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 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 effective measurement depth z* and continue this procedure until all values stabilize. Do this for each sampling/calibration distance (~1, ~33 and ~140 m) separately.
    Equations (5), (6) and (9).
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.
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 ($SOM+B_R$).
15. Average raw neutron counts (N_{raw}) from the moderated sensor (measuring fast neutrons) over 12 h with a moving window.
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).
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}).

18. Correct raw neutron counts for atmospheric pressure variations (N_p).
    Equation (1).
19. Correct raw neutron counts for incoming neutron intensity variations (N_{pi}).
    Equation (2).
20. Correct raw neutron counts for absolute humidity variations (N_{pih}).
    Equation (3).
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).
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 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.

881 Western, A., Zhou, S.-L., Grayson, R., McMahon, T., Blöschl, G., and Wilson, D.: Spatial
882 correlation of soil moisture in small catchments and its relationship to dominant spatial
883 hydrological processes, J. Hydrol., 286, 113-134, doi:10.1016/j.jhydrol.2003.09.014, 2004.
884
885 Zreda, M., Desilets, D., Ferré, T., and Scott, R.: Measuring soil moisture content non-invasively
886 at intermediate spatial scale using cosmic-ray neutrons, Geophys. Res. Lett., 35, L21402,
887 doi:10.1029/2008GL035655, 2008.
888
889 Zreda, M., Shuttleworth, W., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.:
890 COSMOS: the COsmic-ray Soil Moisture Observing System, Hydrol. Earth Syst. Sc., 16, 4079–
891 4099, doi:10.5194/hess-16-4079-2012, 2012.
892

Table 1. Fractions of different tree stands in percent within the footprint of the CRS. The total represents a distance weighted average with an exponential decay towards more distant areas (approximating the exponential distance-weighting from Zreda et al. (2008)).

	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

Kommentar [A9]: Ref3: Totals are not weighted by areas! Area 0-50 m is much smaller than that in 50-150 m, which is smaller than 150-300 m.

Kommentar [A10]: Reply: You are right. We should have specified this before: 'The total represents a distance weighted average with an exponential decay towards more distant areas (approximating the exponential distance-weighting from Zreda et al. (2008)).'

897 Table 2. Overview of the four weighting approaches for other than soil moisture effects on the
898 CRS signal.

Approach	1 SDW	2 DSW	3 DDW	4 DDWnl
consideration of depth-specific W_L and SOM+B_R	no	yes	yes	yes
distance depth-weighting	no	no	yes	yes
non-linear depth-weighting	no	no	no	yes

899

900 Table 3. Example of depth weighting (DSW) for an effective measurement depth of z^* = 22.1 cm,
901 a = 0.0903 and b = 1. Calibration campaign date 21 November 2014 (F4). Note the difference in
902 specific weights if only soil water content θ is considered (wt(z,θ)) or if W_L and $SOM+B_R$ is also
903 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

904

z (cm)	$wt(z,\theta)$	$\int_z^{z+5} wt(z,\theta)$	$wt(z,H_p)$	$\int_z^{z+5} wt(z,H_p)$
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**

905

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

Calibration	P (hPa)	p_{v0} (g m^{-3})	N_{raw} (counts h^{-1})	N_p (counts h^{-1})	N_{pi} (counts h^{-1})	N_{pih} (counts h^{-1})	N_0 (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**

917

Calibration	N_{nm} (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. 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 weighting approaches with 10 calibration campaigns each.

Approach	$(N_0)_\mu$ (counts h^{-1})	$(N_0)_\sigma$ (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	13.7	0.139	0.012
4 DDWnl	828.1	13.3	0.134	0.012

923 Table 6. Modified calibration parameters for the four 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	810.7	0.326	0.001	0.310
4 DDWnl	779.3	0.314	0.001	0.285

924

925 Table 7. Performance measures for the four weighting approaches – comparison of modified
926 calibration (mdf) with standard calibration (stan). KGE' is the modified Kling-Gupta efficiency,
927 β is the bias ratio and γ is the variability ratio. $(KGE')_{\mu}$ and $(KGE')_{\sigma}$ represent the mean and
928 standard deviation of the KGE' values of the 10 individual single-point standard calibrations.

	KGE' mdf	**β mdf**	**γ mdf**	**$(KGE' stan)_{\mu}$**	**$(KGE' stan)_{\sigma}$**	**$(\beta stan)_{\mu}$**	**$(\gamma stan)_{\mu}$**
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	1.076	0.986	0.712	0.081	0.878	1.237
4 DDWnl	0.833	1.148	1.011	0.681	0.096	0.818	1.244

929

930 Table 8. Hydrogen pools (in kg hydrogen per m^2) in the CRS footprint for different moisture
931 conditions (wet: 0.29 m^3 m^{-3}, full canopy and litter storage; intermediate: 0.17 m^3 m^{-3}, dry canopy
932 and moist litter storage; dry: 0.05 m^3 m^{-3}). Above-ground biomass is split into a static part (AGB
933 wet static) comprising stem, branches and dry litter and a variable part (AGB wet variable) that
934 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

935

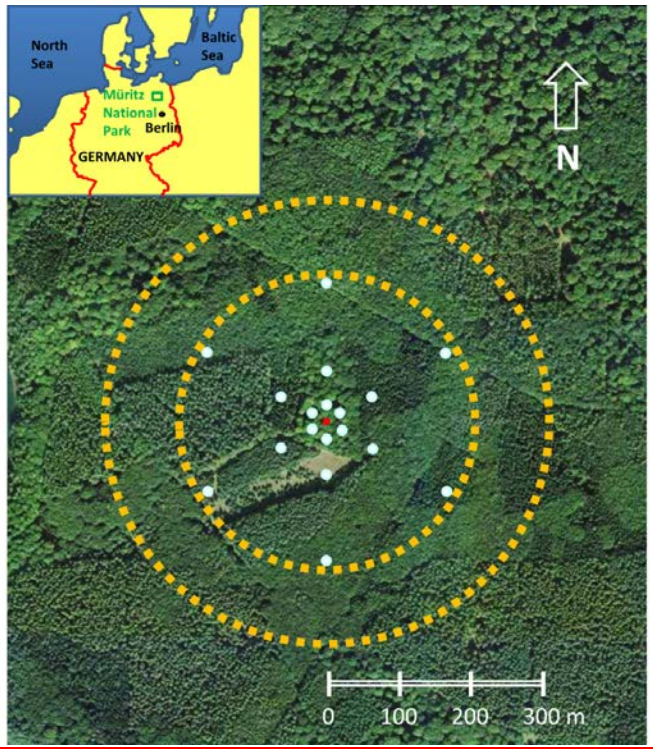


Figure 1. Soil sampling locations for calibration (white~~blue~~ dots) and forest vegetation around the CRS (red dot in the center). The TDT soil moisture sensors are located in close vicinity to the sampling locations. The larger yellow circle approximates the footprint of the CRS as it was assumed 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 [A11]: Ref1: The blue dots are difficult to discern

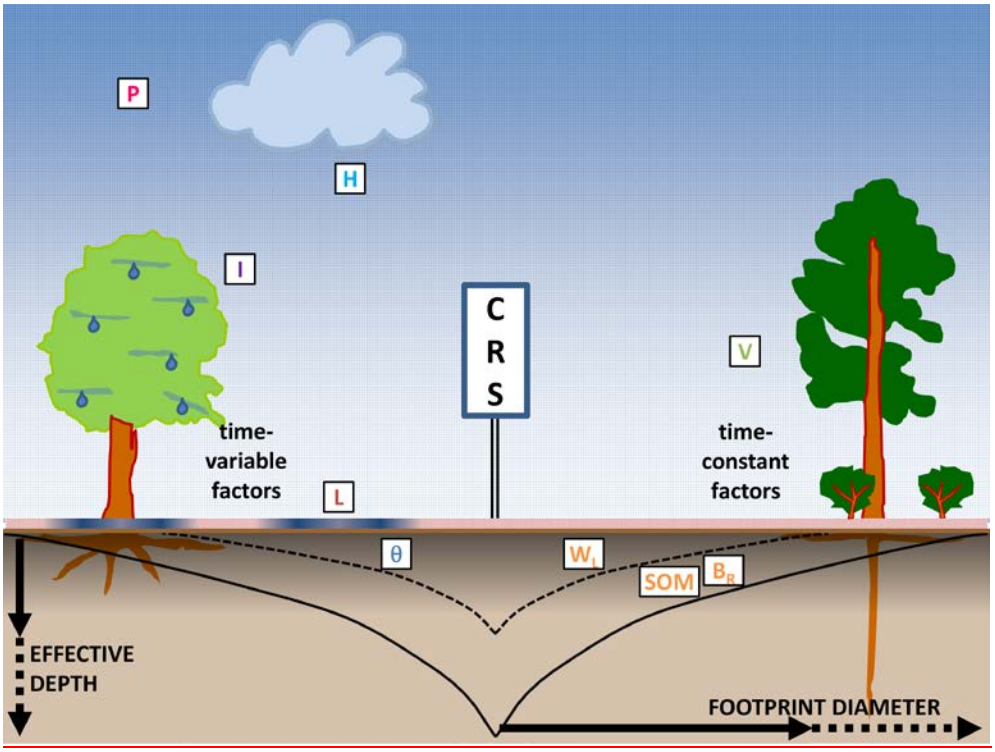


Figure 2. Simplified representation of factors influencing the raw neutron count ~~(N_{raw})~~ and the measurement support volume of the CRS in terms of effective measurement depth and footprint. Temporally variable factors are shown on the left: barometric pressure (P), canopy interception (I), air 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 reduction of the maximum effective depth and footprint diameter (solid black line) and creates a site specific temporally variable effective measurement depth and footprint diameter (dashed black line).

Kommentar [A12]: Ref1: The modified schematic shown in this figure is still confusing. Why should the factors influencing the cosmic-ray intensity a ground level be symbolized a rays of different length? I find this irritating and not helpful for the understanding of the involved processes. In addition, the presented sensing depth boundaries are too numerous be discernible and the distribution is not correct. A better representation can be found in Köhli et al., 2015.Thus, in my opinion the figure should be omitted.

Kommentar [A13]: Reply: We modified the figure to address your concerns. We removed the rays of different length and instead used gradually less intense background color to represent the decreasing number of fast neutrons due to the various factors. We removed all depth boundaries only keeping the maximum extent and one example of reduced depth and footprint. We also modified the shape of the depth boundary so that it resembles better what Köhli et al. (2015) describe.

Kommentar [A14]: Ref3: support volume may be a better term here

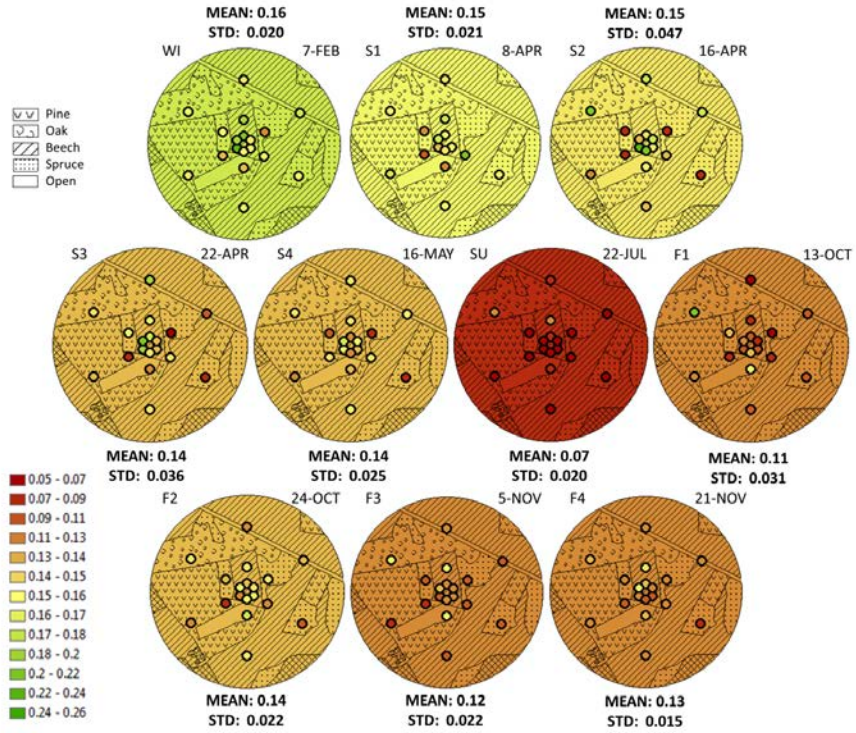
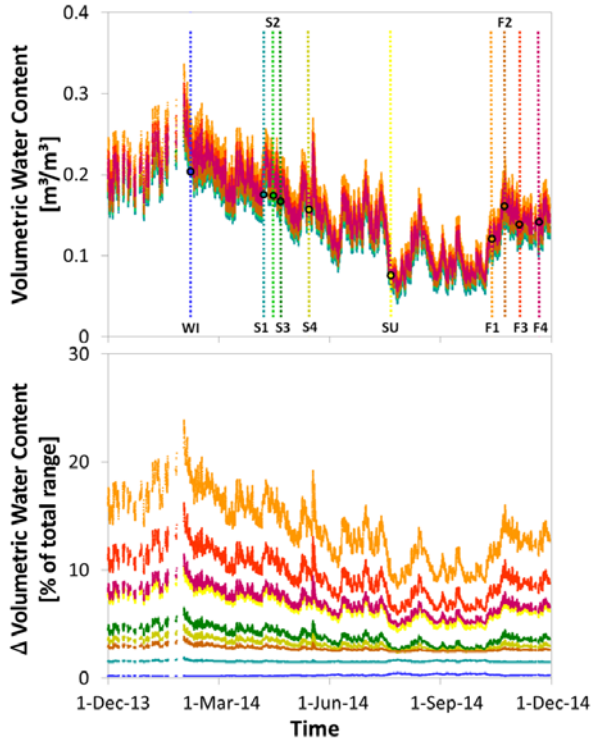


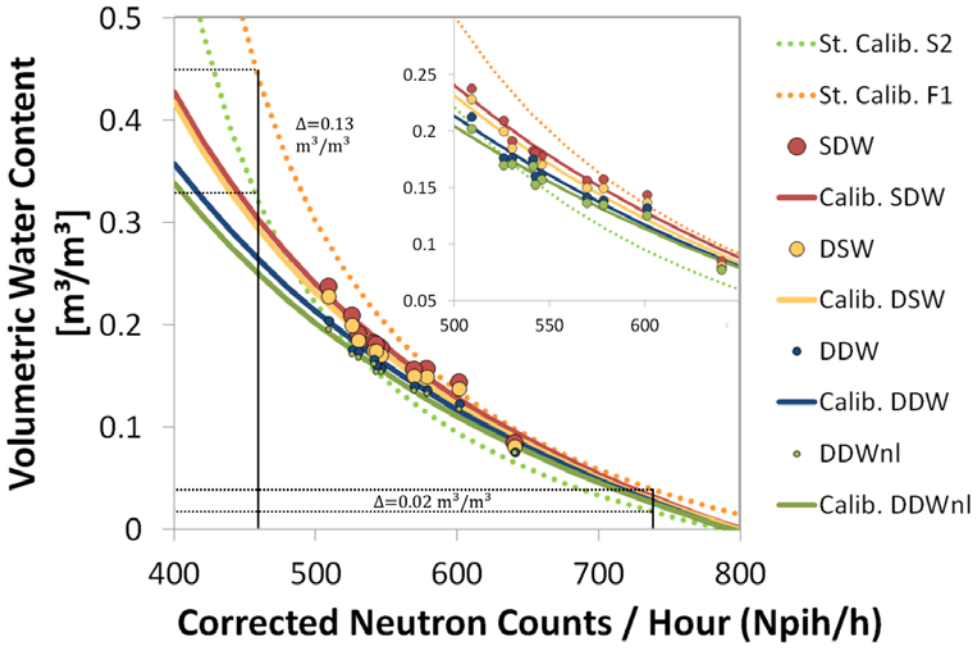
Figure 3. Gravimetrically determined volumetric soil water content patterns in the footprint of the CRS for the 10 calibration dates. The colored dots indicate the unweighted average value from 0 to 30 cm at the 18 calibration locations. Background colors represent the unweighted average value of all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



966

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

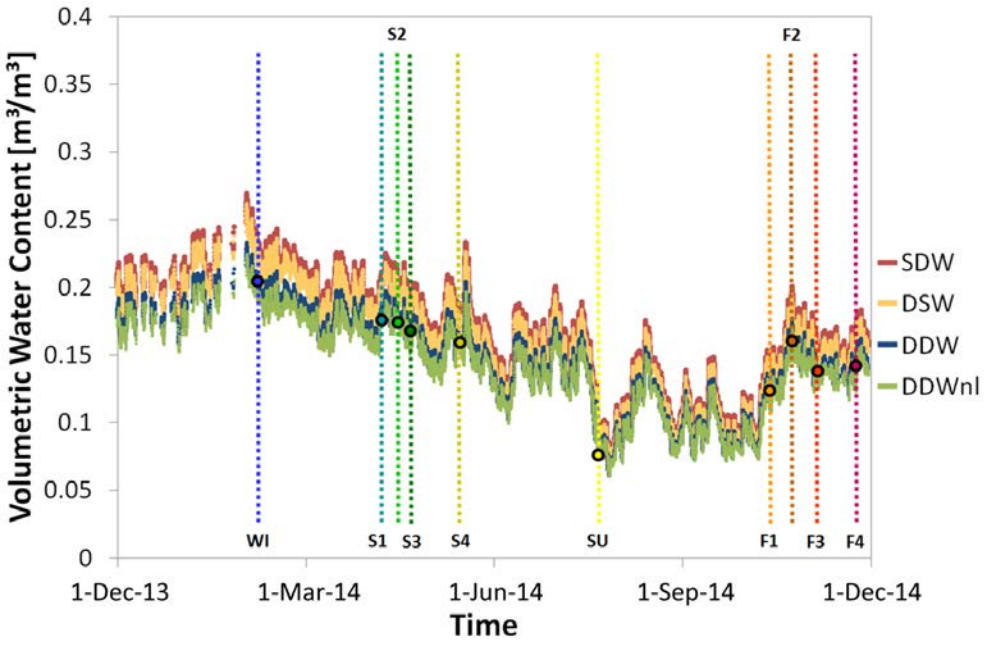
974



975

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

984



985

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

991

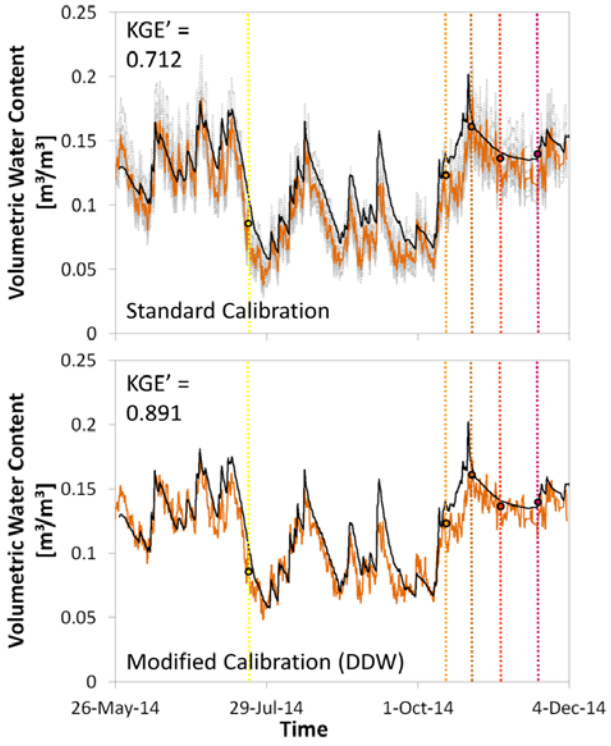


Figure 7. Average volumetric water content derived from TDT point measurements (black line) and CRS measurements (orange line) using different calibration functions. Upper panel: the orange line is an average of the volumetric water content derived from the 10 calibration campaigns of the CRS using the 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.

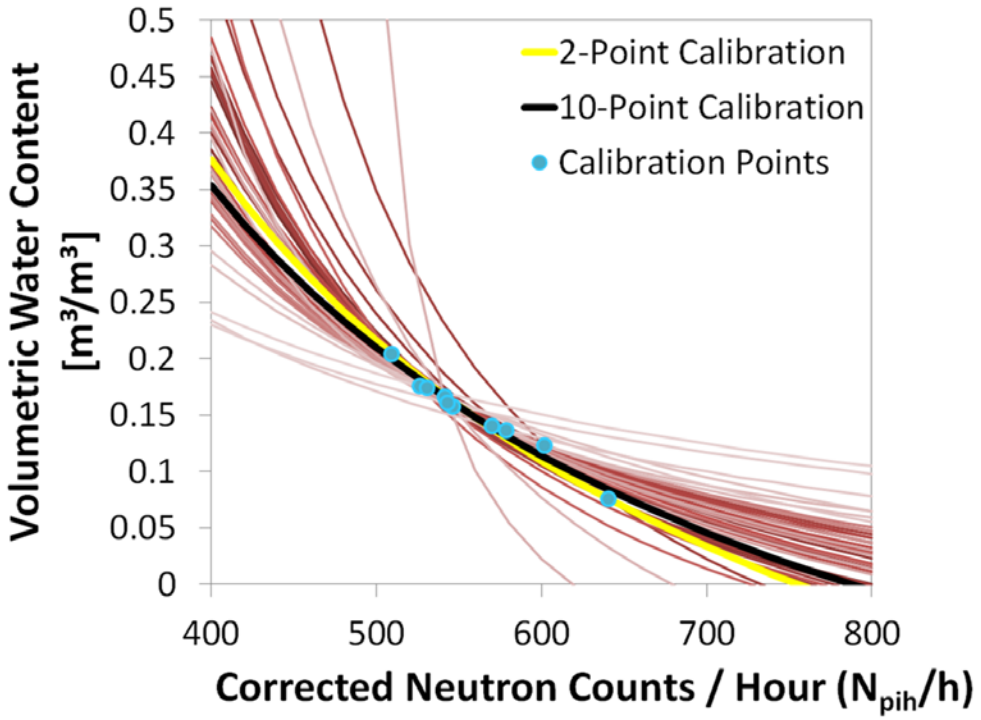


Figure 8. Best-fit N_0-calibration functions (red-brown colored lines) for all combinations of two-point calibrations (blue dots). Best-fit N_0-calibration function for 10-point calibration (black line). Best-fit two-point N_0-calibration function derived from calibration points with highest and lowest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

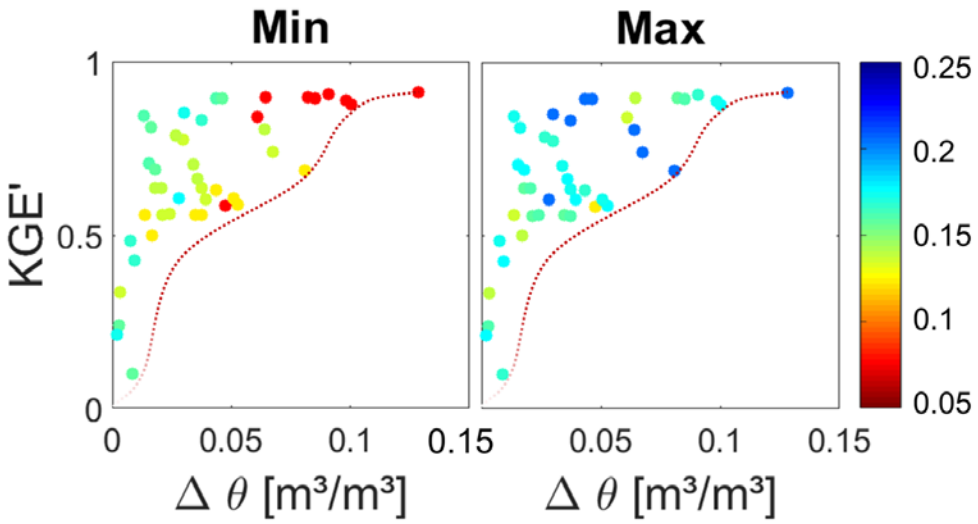


Figure 9. Performance of CRS soil water content data derived from two-point calibrations in relation to difference between soil moisture states ($\Delta\theta$) at the two calibration dates. The color bar indicates volumetric soil water content. Left panel: points are colored according to the soil water 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.

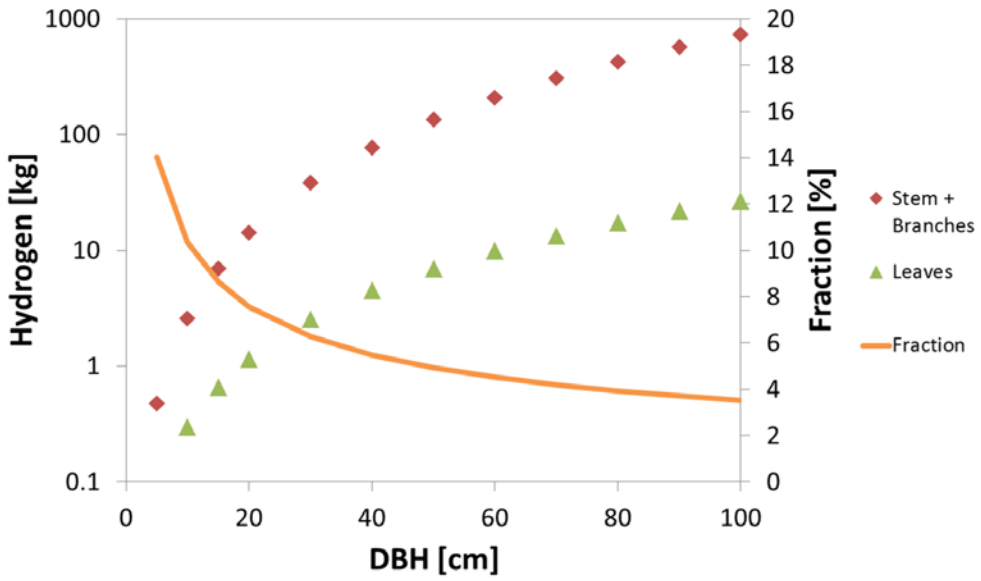


Figure 10. Mass of hydrogen in individual beech trees in stem and branches (red diamonds) and leaves (green triangles) in relation to diameter at breast height (DBH). Fraction of leaf hydrogen mass to total aboveground tree hydrogen mass (orange line).

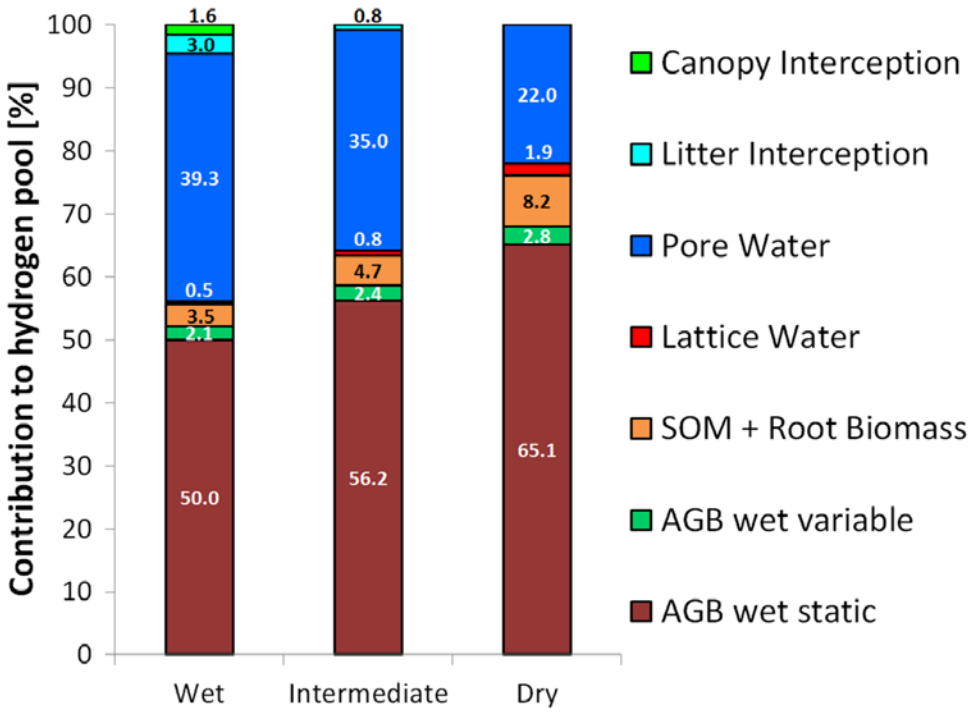


Figure 11: Varying hydrogen pools in the beech forest surrounding the CRS for three different site conditions. AGB (above-ground biomass) wet variable represents hydrogen contained in deciduous leaves (both in the biomass and in the leaf water). AGB wet static comprises hydrogen contained in biomass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

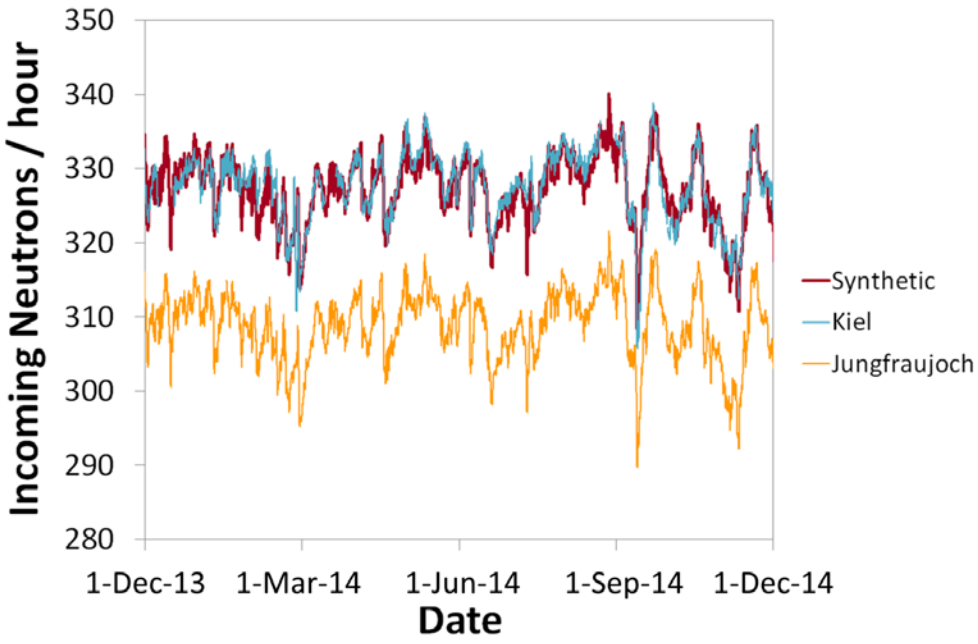
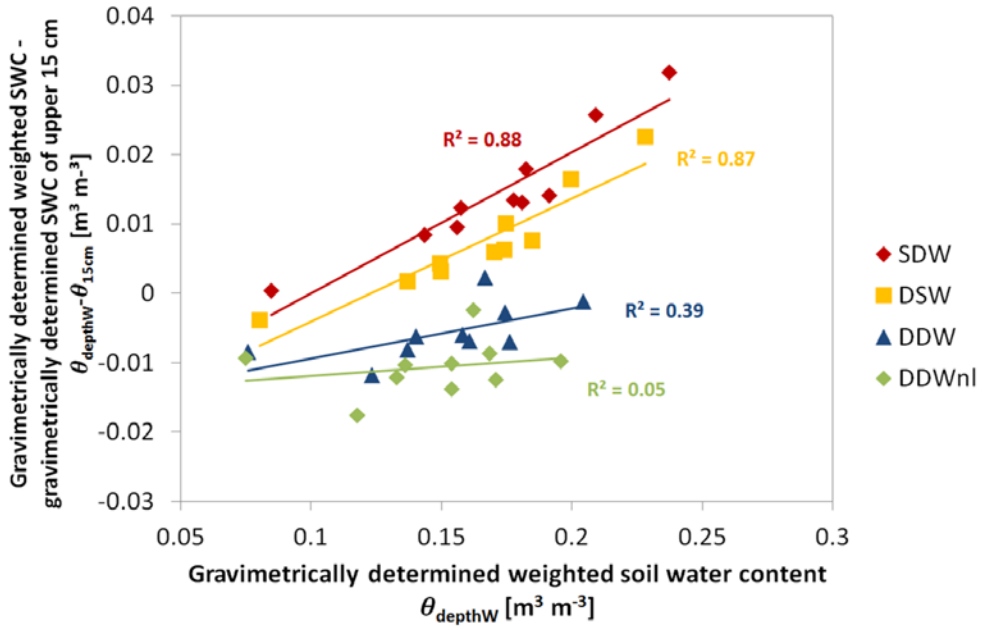


Figure S1. Incoming neutron flux from the neutron monitors in Kiel, Germany and Jungfraujoch, Switzerland and synthetic continuous time series of incoming neutron flux combined from these two and used for the corrections in this study.



1035

1036 Figure S2. Comparison of depth-(and distance-) weighted averages of gravimetrically determined
1037 soil water content with unweighted gravimetrically determined soil water content of the upper 15
1038 cm of the soil. The first two weighting approaches overestimate soil water content in the upper 15
1039 cm especially at high soil water contents. The last two approaches have only a slight negative
1040 offset and no significant relationship with wetness conditions.