Incorporation of globally available datasets into the roving cosmic-ray neutron probe

method for estimating field scale soil water content

William Alexander Avery[1], Catherine Finkenbiner[1], Trenton E. Franz[1], Tiejun Wang[1], Anthony

L. Nguy-Robertson[1], Andrew Suyker[1], Timothy Arkebauer[1,2], and Francisco Munoz-Arriola[1,3]

[1]School of Natural Resources, University of Nebraska-Lincoln

[2]Department of Agronomy and Horticulture, University of Nebraska-Lincoln

[3]Biological Systems Engineering, University of Nebraska-Lincoln

Keywords: Cosmic-ray neutron probe; soil moisture; calibration parameters; remote sensing;

maize; soybean

Corresponding author T.E. Franz (tfranz2@unl.edu)

Abstract

The need for accurate, real-time, reliable, and multi-scale soil water content (SWC)

monitoring is critical for a multitude of scientific disciplines trying to understand and predict the

earth's terrestrial energy, water, and nutrient cycles. One promising technique to help meet this

demand is fixed and roving cosmic-ray neutron probes (CRNP). However, the relationship

between observed low-energy neutrons and SWC is affected by local soil and vegetation

calibration parameters. This effect may be accounted for by a calibration equation based on local

21   soil type and the amount of vegetation. However, determining the calibration parameters for this

22   equation is labor and time intensive, thus limiting the full potential of the roving CRNP in large

23   surveys and long transects, or its use in novel environments. In this work, our objective is to

24   develop and test the accuracy of globally available datasets (clay weight percent, soil bulk

25   density, and soil organic carbon) to support the operability of the roving CRNP. Here, we

26   develop a 1 km product of soil lattice water over the CONtinental United States (CONUS) using

27   a database of *in-situ* calibration samples and globally available soil taxonomy and soil texture

28   data. We then test the accuracy of the global dataset in the CONUS using comparisons from 61

29   *in-situ* samples of clay percent (RMSE = 5.45 wt. %, R^2 = 0.68), soil bulk density (RMSE =

30   0.173 g/cm^3, R^2 = 0.203), and soil organic carbon (RMSE = 1.47 wt. %, R^2 = 0.175). Next, we

31   conduct an uncertainty analysis of the global soil calibration parameters using a Monte Carlo

32   error propagation analysis (maximum RSME ~0.035 cm^3/cm^3 at a SWC = 0.40 cm^3/cm^3). In

33   terms of vegetation, fast growing crops (i.e. maize and soybeans), grasslands, and forests

34   contribute to the CRNP signal primarily through the water within their biomass and this signal

35   must be accounted for accurate estimation of SWC. We estimated the biomass water signal by

36   using a vegetation index derived from MODIS imagery as a proxy for standing wet biomass

37   (RMSE < 1 kg/m^2). Lastly, we make recommendations on the design and validation of future

38   roving CRNP experiments.

39

40   **1. Introduction**

41   By the year 2050, over nine billion people are predicted to inhabit the Earth (United

42   Nations, 2015). The monumental task of feeding the projected global population will require a

43   near doubling of grain production (FAO, 2009). As of today, the majority (~2/3) of water

44 consumption by humans is used for agriculture, where approximately half of all global food

45 production comes from irrigated agriculture (Mekonnen et al., 2011). As such, an increase in

46 food demand will put an even greater demand on fresh water resources, particularly an

47 increasing reliance on groundwater (Mekonnen et al., 2011). The ability to model and forecast

48 the hydrologic cycle will continue to play a major role in effective water resource management

49 in the coming decades. Currently, most land surface models (LSM) aimed at characterizing the

50 fluxes of water, energy, and nutrients, have relied on either sparse point scale *SWC* monitoring

51 networks (Crow et al. 2012) or remote sensing products with large pixel sizes (~36 km) and

52 shallow penetration depths (Kerr et al., 2010 and Entekhabi et al., 2010). A critical scale gap

53 exists between these methods requiring innovative monitoring strategies (Robinson et al., 2008).

54 Moreover, as LSMs continue to move towards highly refined spatial resolutions of 1 km or less

55 (Wood et al., 2011), the need for accurate and spatially exhaustive *SWC* datasets continues to

56 grow (Beven and Cloke, 2012).

57   Estimating and monitoring *SWC* at the appropriate spatial and temporal scale for effective

58 incorporation into LSMs has proven to be a difficult task. On one hand, monitoring networks at

59 the regional (e.g., Nebraska Automated Weather Data Network; AWDN, Oklahoma Mesonet)

60 and continental scales (Climate Reference Network; CRN, Soil Climate Analysis Network;

61 SCAN) have continuously recording point sensors. However, these sparse networks are difficult

62 to place in the context of the surrounding landscape given the multifractal behavior that soil

63 moisture fields exhibit (Korres et al. 2015). Techniques such as temporal stability analysis

64 (Vachaud et al., 1985) can help improve the representativeness of the monitoring networks but

65 require *a priori* spatial information. On the other hand, remote sensing satellites using passive

66 microwaves can monitor global *SWC* data every few days albeit with large spatial footprints (~36

67 by 36 km, Entekhabi et al., 2010 and Kerr et al., 2010). In addition, passive microwaves lack

68 significant penetration depths (~ 2-5 cm Njoku et al., 1996), limiting their effectiveness as a

69 remote sensing input for full root zone coverage in LSMs.

70   Alternatively, the field of geophysics offers a variety of techniques to help fill the spatial

71 and temporal gaps between point sensors and remote sensing products (Bogena et al., 2015).

72 Bridging this gap requires both novel geophysical techniques and integrated modeling strategies

73 capable of merging both point and remotely sensed data into a unified framework (Binley et al.,

74 2015). One promising geophysical technique to help fill this need is fixed (Desilets et al., 2010,

75 Zreda et al., 2012) and roving cosmic-ray neutron probes (CRNP; Chrisman et al., 2013, Dong et

76 al., 2014), which measure the ambient amount of low-energy neutrons in the air. The low-energy

77 neutrons are highly sensitive to the mass of hydrogen, and thus SWC, in the near surface (Zreda

78 et al., 2012). CRNPs estimate the area-average SWC because neutrons are well mixed within the

79 footprint of the sensor which typically has a radius of several hundred meters and depths of tens

80 of decimeters (Desilets and Zreda 2013, Köhli et al., 2015).

81   To date, the CRNP method has been mostly used as a fixed system in one location to

82 continuously measure SWC as part of a large monitoring network (Zreda et al., 2012, Hawdon et

83 al., 2014). Recent advancements have allowed the CRNP to be used in mobile systems to

84 monitor transects across Hawaii (Desilets et al., 2010), monitor entire basins in southern Arizona

85 (Chrisman et al., 2013), compare against remote sensing products in central Oklahoma (Dong et

86 al., 2014), and monitor ~140 agricultural fields in eastern Nebraska (Franz et al., 2015). In order

87 to accurately estimate SWC, the CRNP method relies on a calibration function to convert

88 observed low-energy neutron counts into SWC (Desilets et al., 2010, Bogena et al., 2013, see

89 Sec. 2.2 for full details). The calibration procedure requires site specific sampling of both soil

90   and vegetation data in order to determine the required parameters. While the calibration of a

91   fixed CRNP is fairly standardized (Zreda et al., 2012; Franz et al., 2012; Iwema et al., 2015,

92   Baatz et al., 2014), the heterogeneous nature of soil and vegetation characteristics across a

93   landscape makes the pragmatic calibration of the roving CRNP a significant challenge.

94   Specifically, the presence of water within vegetation and the soil minerals may alter the shape of

95   the local calibration function and thus accuracy of *SWC* (Baatz et al., 2015). The need for

96   reliable, accurate, depth-dependent, and localized soil and vegetation spatial information for use

97   in the calibration function is critical in order to fully exploit the potential of the roving CRNP to

98   monitor landscape scale *SWC* across the globe.

99   The objective of this study is to explore the utility and accuracy of currently available

100   global soil and vegetation datasets (soil organic carbon, soil bulk density, soil clay weight

101   percent, and crop biomass) for use in the calibration function. To accomplish our objective, we

102   aimed to answer the following questions:

103   1) Can global datasets of soil bulk density, soil organic carbon, and soil clay weight percent be

104   used instead of *in-situ* sampling within reasonable error for use in the roving CRNP calibration

105   function?

106   2) Can the use of remotely sensed vegetation products, specifically the Green Wide Dynamic

107   Range Vegetation Index (GrWDRVI) be used to quantify fresh biomass with reasonably low

108   error (< 1 kg/m^2) for use in the roving CRNP calibration function?

109   To answer these questions, we tested the accuracy of these datasets against *in-situ* sample

110   datasets of the same parameters. Existing *in-situ* datasets from across the CONtinental United

111   States (CONUS) were combined with *in-situ* datasets from eastern Nebraska, which focused on

112   fast growing crops of maize and soybean. Specifically, we tested the accuracy and use of a ~1

113   km global soil dataset (Shangguan et al., 2014). In addition, we examined the use of the Green

114   Wide Dynamic Range Vegetation Index (*GrWDRVI*, Gitelson, 2004) derived from NASA's

115   MODIS sensor aboard the Terra satellite for use in estimating the amount of fresh crop biomass.

116   The remainder of the paper is organized as follows: In the Methods section, the CRNP

117   method is first presented, with emphasis on the integration of the calibration function and soil

118   and vegetation parameters to convert observed low-energy neutron counts into *SWC*. Next, *in-*

119   *situ* methods for estimating the soil and vegetation calibration parameters are discussed, which is

120   followed by discussions on the soil and vegetation products available globally at ~1 km

121   resolution. In the Results section, we first compare the *in-situ* soil sampling against the global

122   datasets. Next, we develop a 1 km CONUS soil lattice water map using *in-situ* samples. We then

123   compare the *GrWDRVI* against *in-situ* samples from Nebraska to estimate the changes in maize

124   and soybean fresh biomass. Lastly, we present an error propagation analysis investigating the

125   potential uncertainty of using the global soil calibration data vs. local *in-situ* sampling. The paper

126   concludes with a discussion on best practice recommendations for calibrating and validating a

127   roving CRNP experiment.

128

129   **2. Methods**

130   2.1 Overview of Cosmic-ray Neutron Probe

131   The CRNP estimates area-averaged *SWC* via measuring the intensity of low-energy

132   neutrons (i.e. ~epithermal) near the ground surface (Zreda et al. 2008, 2012). A cascade of

133   neutrons with a continuous energy spectrum are created in the earth's atmosphere when

134   incoming higher energy particles produced within supernovae interact with atmospheric nuclei

135   (Zreda et al., 2012 and Köhli et al., 2015). After fast neutrons are created, they continue to lose

136   energy during numerous collisions with nuclei in air and soil, and become low-energy neutrons

137   that are detected with the probe. The abundance of hydrogen atoms in the air and soil largely

138   controls the removal rate of low-energy neutrons from the system (Zreda et al. 2012). Water in

139   the near surface soil (i.e. *SWC*) is one of the largest sources of hydrogen present in terrestrial

140   systems (McJannet et al. 2014). Thus, relative changes in the intensity of low-energy neutrons

141   are overwhelmingly due to changes in the *SWC*. However, the shape of the calibration function

142   (see section 2.2) is somewhat modified by local soil and vegetation parameters (Zreda et al.

143   2012) reflecting the variation of background hydrogen levels across landscapes.

144   Using a standard neutron detector with a 2.54 cm layer of plastic, Zreda et al. (2008) first

145   described the support volume the detector measures to be a circle of ~300 m in radius with

146   vertical penetration depths of 12 to 76 cm depending on *SWC*. Recent neutron transport

147   modeling has further refined the footprint area to be a function of atmospheric water vapor,

148   elevation (Desilets and Zreda, 2013), surface heterogeneity (Köhli et al., 2015), vegetation

149   (Köhli et al., 2015), and *SWC* (Köhli et al., 2015). Köhli et al. (2015) found the footprint to range

150   between 130 and 240 m in radius depending on conditions. Despite the varying footprint

151   characteristics, the large measurement area of tens of hectares makes this non-invasive technique

152   an ideal complement to long-term surface energy balance monitoring around the globe.

153   Currently, there are >200 fixed CRNP (personal communication with Darin Desilets of

154   HydroInnova LLC, Albuquerque, NM) functioning in this capacity around the United States of

155   America (Zreda et al., 2012), Australia (Hawdon et al., 2014), Germany (Baatz et al., 2014),

156   South Africa, China, and the United Kingdom. The real-time *SWC* data provide critical

157   infrastructure for use in weather forecasting and data assimilation in LSMs (Shuttleworth et al.,

158   2013, Rosolem et al., 2014, Renzullo et al., 2014).

159   In addition to the fixed CRNP measuring hourly *SWC*, a roving version of the CRNP has

160   been used to reliably measure *SWC* at temporal resolutions as low as 1 minute (Chrisman et al.,

161   2013; Dong et al., 2014) providing the ability to make *SWC* maps over hundreds of square

162   kilometers in a single day. Moreover, Franz et al. (2015) found that a combination of fixed and

163   roving CRNP data in a statistical framework has the ability to form an accurate, real-time, and

164   multiscale monitoring network. With the continued increase in observation spatial scales, the use

165   of *in-situ* sampling in the traditional CRNP calibration procedure is no longer practical, thus

166   requiring the use of alternative available datasets to improve its operability. The remainder of

167   this work will first describe the availability of such global datasets and then test the accuracy of

168   using the datasets in the CNRP calibration function.

169

170   **2.2 The Cosmic-ray Neutron Probe Calibration Function**

171   In order to convert observed low-energy neutron measurements into *SWC*, a series of

172   scaling factors, correction factors, and calibration functions have been developed. Zreda (2012)

173   describes in detail the affects from changes in geomagnetic latitude, changes in incoming high-

174   energy cosmic-ray intensity, and atmospheric pressure. Rosolem et al. (2013) further describes

175   changes in absolute air humidity near the surface. Following these four scaling and correction

176   factors, the corrected low-energy neutron counts can be converted into *SWC*. Desilets et al.

177   (2010) proposed the original calibration function (Eq. 1) valid for mass based gravimetric

178   measurements which Bogena et al. (2013) further expanded for volumetric water content. The

179　calibration function has been successfully tested against direct sampling and point sensor

180　measurements with RMSE < 0.03 cm^3/cm^3 across the globe including arid shrublands in

181　Arizona, USA (Franz et al., 2012), semi-arid forests in Utah, USA (Lv et al., 2014), to humid

182　forests in Germany (Bogena et al., 2013), and across ecosystems in Australia (Hawdon et al.,

183　2014). The original calibration function proposed by Desilets et al., (2010) is:

184　$$\theta_T = \left(\frac{a_0}{\frac{N}{N_0} - a_1} - a_2 \right) \qquad (1)$$

185　where θ_T (g/g) is the total gravimetric water content, $a_0 = 0.0808$, $a_1 = 0.3720$, $a_2 = 0.1150$ (see

186　Desilets et al., (2010) for details), N(counts per time interval) is the aforementioned low-energy

187　corrected neutron count rate, and N_0 (counts per time interval) is the theoretical counting rate at a

188　location with dry silica soils. Zreda et al. (2012) illustrated that:

189　$$\theta_T = \theta_p + \theta_{LW} + \theta_{SOC} \qquad (2)$$

190　where θ_p (g/g) is the gravimetric pore water content in the soil, θ_{LW} (g/g) is the soil lattice water,

191　and θ_{SOC} (g/g) is the soil organic carbon water equivalent. The volumetric soil water content,

192　SWC, (cm^3/cm^3) is found by multiplying θ_p by $\frac{\rho_b}{\rho_w}$, where ρ_b (g/cm^3) is dry soil bulk density and

193　$\rho_w = 1$ g/cm^3 is the density of water.

194　　　　To account for effects of time varying above-ground vegetation on the low-energy

195　neutron counts (Franz et al., 2013; Coopersmith et al., 2014), Franz et al. (2015) proposed the

196　following additional correction factor to N_0:

197　$$N_0(BWE) = m * BWE + N_0(0) \qquad (3)$$

198   where $N_0(0)$ is the instrument specific estimate of N_0 with no standing biomass, BWE is the

199   biomass water equivalent (kg/m^2 ~ mm of water/m^2), and m is the slope of the relationship

200   between N_0 and BWE, determined via *in-situ* calibration datasets. The BWE is further defined

201   as:

202    $$BWE = SWB - SDB + SDB * f_{WE} \qquad\qquad (4)$$

203   where SWB is the standing wet biomass per unit area (kg/m^2 ~ mm of water/m^2), SDB is the

204   standing dry biomass per unit area (kg/m^2 ~ mm of water/m^2), and $f_{WE} = 0.494$ is the

205   stoichiometric ratio of H_2O to organic carbon (assuming organic carbon is cellulose, $C_6H_{10}O_5$).

206   Using nine *in-situ* calibration datasets for maize and soybean crops, Franz et al. (2015) found

207   their roving CRNP had a statistically significant linear relationship between N_0 and BWE

208   yielding $N_0(0) = 518.34$ counts per minute and $m = -4.9506$ (R^2 = 0.515 and p-value = 0.03).

209   We note the coefficients are less suitable for forest canopies given the need for a neutron

210   geometric efficiency factor described further in the supplemental material of Franz et al. (2013).

211   We also refer the reader to Coopersmith et al. (2014) and Baatz et al. (2015) for further

212   discussion of CRNP use in forest canopies, and Bogena et al. (2013) for a discussion of below-

213   ground biomass and litter layers. In addition, plant specific root-shoot ratios (Peichl et al., 2012)

214   or allometric relationships (Jenkins et al., 2003) may be used to derive a better understanding of

215   the impact of time-varying below-ground biomass on N_0. This is an open and challenging

216   research area and beyond the scope of the current work.

217

218   **2.3 *In-situ* Soil and Vegetation Calibration Parameters**

219   In the simplest form, the calibration function summarized in equations (1-4) requires

220   depth-average estimates of three soil parameters, θ_{LW}, θ_{SOC}, and ρ_b, and two vegetation

221   parameters SWB and SDB. We note that depth-weighted average parameters, belowground

222   biomass, and depth-weighted SWC are needed to fully understand the decreasing sensitivity of

223   the CRNP with depth as recommended elsewhere (Bogena et al., 2013 and Köhli et al., 2015).

224   As a first step, here we will only consider depth and area-average properties given the resolution

225   of the global remote sensing products. We expect future work to improve on these analyses as

226   regional datasets contain higher spatial resolution data. In order to estimate depth and area-

227   average soil parameters, Zreda et al. (2012) and Franz et al. (2012) recommended averaging 108

228   individual *in-situ* soil samples from 18 locations (every 60 degrees and radii of 25, 75, 200 m)

229   and six depths (every 5 cm from 0-30 cm) within a CRNP footprint. In light of recent modeling

230   work (Köhli et al. 2015), this sampling pattern may need to be adjusted to be more representative

231   of encountered conditions (such as shorter sampling distances due to reduced footprint area).

232   Given the mixture of previously published datasets and new datasets used here, we decided to

233   use the original sampling location description. Zreda et al. (2012) found that a composite sample

234   of 1 g of material gathered from each of the 108 samples was adequate to estimate θ_{LW} and θ_{SOC}.

235   These composite samples can be analyzed directly for lattice water (g/g), soil total carbon (TC,

236   g/g), and inorganic carbon (TIC, g/g) determined by measuring CO_2 after the sample is acidified

237   (e.g. by Actlabs of Ontario Canada, Analysis Codes: 4E-exploration, 4F-CO2, 4F-C, and 4F-

238   H2O+/-). Franz et al. (2015) reported $\theta_{SOC} = (TC - TIC) * 1.724 * f_{WE}$, where 1.724 is a

239   constant to convert total organic carbon into total organic matter and f_{WE} is given above. To

240   estimate ρ_b at each location, Zreda et al. (2012) used a 30 cm long split tube auger, which

241 contained six 5 cm diameter by 5 cm length rings. All samples were then averaged to get a

242 composite value.

243 In order to estimate standing wet biomass (*SWB*) and standing dry biomass (*SDB*) in

244 maize and soybeans, Franz et al. (2015) measured average plant density in 1 m^2 quadrats at each

245 of the 18 sampling locations. In a subset of six sites (randomly chosen from one radius for each

246 of the six transects) three plants were removed and placed in a paper bag for weighing within

247 two hours (to minimize water loss). The plants were then dried for five days at 70°C and

248 weighed again. Using the density of plants, wet weight, and dry weight, *SWB* and *SDB* can be

249 determined at each site and averaged across the CRNP footprint.

250

251 **2.4 Global Datasets of Soil Properties**

252 Shangguan et al. (2014) compiled a thirty arc second (~1 km) Global Soil Dataset

253 (GSDE) with 34 soil parameters in 8 layers (0–0.045, 0.045–0.091, 0.091–0.166, 0.166–0.289,

254 0.289–0.493, 0.493–0.829, 0.829–1.383, and 1.383–2.296 m). In order to construct an average

255 value relevant to the CRNP, we arithmetically averaged the top four layers in each grid location

256 to form a composite value (~30 cm) over the CONUS. The GSDE contains estimates of soil bulk

257 density and soil organic carbon. In order to construct a map of lattice water, we explored if any

258 relationships existed between clay weight fraction and lattice water following the work of

259 Greacen et al. (1981) using active neutron probe calibration procedures developed for Australian

260 soils. In order to account for variations in chemical and physical weathering on lattice water

261 (Zreda et al., 2012), we further partitioned the analyses based on soil order. A global soil order

262 map with a resolution of five arc minutes (~ 8 km) containing 25 major soil classifications was

263 first uploaded to ArcMap (ESRI, v. 10.2.2) and clipped to the CONUS. The 25 soil

264 classifications were then categorized into 12 major classifications of U.S. soil taxonomy (see Fig.

265 1, personal communication with Prof. M. Kuzila, University of Nebraska-Lincoln, Soil Survey

266 Staff, 1999). The reduction from 25 to 12 soil classifications allowed us to generate larger

267 sample sizes for each classification from the available calibration datasets. Using the available

268 lattice water samples from Zreda et al. (2012) and additional samples collected *in-situ* over 2014,

269 we analyzed if any statistically significant relationships existed between GSDE clay weight

270 percent and 61 *in-situ* lattice water samples for each of the US soil orders (Table S1). We note

271 that this procedure could be used globally if *in-situ* lattice water samples were available for all 25

272 soil taxonomic groups. From these relationships, a map of the CONUS lattice water weight

273 percent was developed by using either the mean value of the *in-situ* lattice water or the linear

274 relationships between clay weight percent (from the GSDE) and the lattice water *in-situ* samples.

275 A statistically significant p value (<0.05) was used to discriminate between using the mean

276 values and linear relationship. Additionally, *in-situ* samples of soil organic carbon, bulk density,

277 and clay weight percent were compared against the same parameters derived from the GSDE.

278

279 **2.5 Global Datasets of Vegetation Properties**

280   In order to estimate *SWB* and *SDB*, we downloaded remotely sensed 500 m MODIS

281 reflectance data from NASA's Terra satellite (http://earthexplorer.usgs.gov/). To calibrate and

282 validate the *in-situ* vegetation data to the remotely sensed vegetation estimates, we sampled two

283 different agricultural areas in eastern Nebraska. The MODIS reflectance data were used to

284 generate a widely used vegetation index (see detailed information below), and then calibrated

285 against historical biomass data (2003-2013) from 3 fields near Mead, NE. Each field is part of

286   the AmeriFlux network (http://ameriflux.ornl.gov/) with data going back to 2001 (site

287   description given in Suyker et al., 2005). Each field is approximately 65 ha in area. Field 1

288   (Mead Irrigated/US-Ne1, 41.1650°, -96.4766°) is irrigated with continuous maize. Field 2 (Mead

289   Irrigated Rotation/US-Ne2, 41.1649°, -96.4701°) is irrigated with a rotation of maize and

290   soybean. Field 3 (Mead Rainfed/US-Ne3, 41.1797°, -96.4396°) is rainfed with a rotation of

291   maize and soybean. At these three fields, destructive biomass samples were collected

292   approximately every two weeks at 6 different locations in the field, typically consisting of 30-35

293   individual plants per sampling bout. From the destructive sampling bouts, we were able to

294   compute *SWB* and *SDB*. The sites, with their long sampling records consisting of both rainfed

295   and irrigated soybean and maize, are an ideal location for calibrating the remote sensing

296   reflectance data and vegetation indices. In order to validate the derived vegetation index and

297   coefficients from the above mentioned three sites, we used 4 bouts of destructive biomass

298   sampling at two fields (each approx. 65 ha.) during 2014 near Waco, NE (Franz et al. 2015). The

299   fields were irrigated maize (40.9482°, -97.4875°) and irrigated soybean (40.9338°, -97.4587°).

300   *SWB* and *SDB* were collected following the protocol described in section 2.3.

301   A total of 924 MODIS images over the growing seasons (May to October) between 2003

302   and 2014 were downloaded for calibration and validation of the corresponding destructive

303   biomass samples at the five field sites in central and eastern Nebraska (note: MODIS images

304   from the closest date to *in-situ* sampling were used with up to a 4 day offset). We extracted the

305   MODIS reflectance data in the green and near-infrared electromagnetic spectrum range. Next,

306   we removed any pixels that were skewed by incidental cloud cover (Nguy-Robertson & Gitelson,

307   2015). The resulting data were then transformed from separate reflectance images into the Green

308   Wide Dynamic Range Vegetation Index (*GrWDRVI*; Gietelson, 2004):

309 $$GrWDRVI = \frac{(0.1*Near\ Infrared - Green)}{(0.1*Near\ Infrared + Green)}$$ (5)

310 where near-infrared light (MODIS band 2) has wavelength between 841 and 876 nm and green

311 light (MODIS band 4) has wavelength between 545 and 565 nm. The *GrWDRVI* has been shown

312 to have better correlations with observed *in-situ* biomass as compared to other vegetation indices

313 such as NDVI (Nguy-Robertson et al., 2012; Nguy-Robertson & Gitelson, 2015). We then

314 investigated if any relationships existed between *GrWDRVI* and *SWB* and *SDB*. We note that a

315 variety of vegetation indices exist in the literature (c.f. Kumar et al. 2015 and Duncan et al.

316 2015) and that this analysis is a first step for use with maize and soybean. We anticipate that

317 other vegetation indices may be more appropriate with use in other crops or vegetation types and

318 more research is needed in this area.

319

320 **2.6 Error Propagation Analysis of GSDE Soil Properties**

321   We used a Monte Carlo analysis to estimate the expected uncertainty if the GSDE

322 parameters were used instead of *in-situ* estimates. The statistical metrics of root mean square

323 error (RMSE), mean absolute error (MAE), and bias describe the error propagation in the Monte

324 Carlo simulation experiment. From the 61 CONUS *in-situ* samples and the GSDE soil properties,

325 we estimated the mean difference and the covariance matrix for θ_{LW}, θ_{SOC}, and ρ_b. With these

326 data, we simulated 100,000 realizations of the "true" (i.e. from the *in-situ* sampling) and

327 perturbed soil properties using a multivariate normal distribution. Using a range of observed

328 neutron counts and solving equations (1-2) with the true and perturbed soil properties, we also

329 estimated the true and perturbed *SWC*. In order to provide realistic constraints on the error

330 propagation results, we assumed soil bulk density was constrained between 1.2-1.5 g/cm^3, lattice

15

331   water between 1-8 wt. %, soil organic carbon between 0-8 wt. %, and *SWC* between 0.03-0.45

332   cm^3/cm^3. Simulated and calculated values outside of these bounds were either reset to the

333   minimum or maximum value or removed from the Monte Carlo statistics. A minimum threshold

334   of 70% of simulated cases was used to compute all error statistics for each case. We note that the

335   effects of growing biomass were not included here given the lack of available calibration datasets

336   at all sites, but could be incorporated in future work following a similar methodology.

337

338   **3. Results**

339   **3.1. Comparison of *In-situ* and Global Soil Calibration Parameters**

340   The comparisons between observed clay weight percent, soil bulk density, soil organic

341   carbon and the GSDE values are summarized in Table S1 and Figure 2 a, b, c for the 61

342   sampling sites within the CONUS. Other than 1 outlier (see discussion in 4.1.), the comparison

343   between the mean observed and GSDE clay weight percent (of sites that had clay weight

344   percent) behaved well (RMSE = 5.45 wt. %, $R^2 = 0.68$) considering the difference in scale and

345   methods. The comparisons between soil bulk density (RMSE = 0.173 g/cm^3, $R^2 = 0.203$) and soil

346   organic carbon as it was during the various 2011-2014 sampling campaigns, (RMSE = 1.47 wt.

347   %, $R^2 = 0.175$) generally followed the same positive trend. We note that the slope of the

348   relationships for soil bulk density and soil organic carbon is different from 1 and can lead to

349   biased results. Caution should be used for using these estimates as opposed to local in-situ

350   sampling.

351   In order to construct a map of the CONUS lattice water, we investigated if any significant

352   relationships existed between GSDE clay wt. % and observed lattice water for each US soil

353 taxonomic group (Table 1). We found that a significant linear relationship existed between clay

354 wt. % and lattice water for all 61 sites (R^2 = 0.183, p value <0.001). However, after partitioning

355 the sites into soil taxonomic groups, only the mollisol taxonomic group yielded a statistically

356 significant relationship (R^2 = 0.539, p value <0.001). Therefore, in order to construct a CONUS

357 lattice water map, we used the mean values for six taxonomic groups and neglected the

358 remaining five taxonomic groups due to an inadequate number of samples (Figure 3). Figure 2d

359 illustrates the comparison between the derived and observed lattice water for the 61 CONUS

360 sites (RMSE = 1.299 wt. %, R^2 = 0.315). Table S1 summarizes the observed and GSDE values

361 for all 61 sites and Table 2 summarizes the mean difference and covariance matrix between the

362 *in-situ* values and GSDE values. The mean difference and covariance differences were used in

363 the error propagation analysis described in section 2.6 and 3.3. We note that each of the mean

364 differences followed a normal distribution (see Table S1 for *in-situ* and GSDE values).

365

366 3.**2. Comparison of *In-situ* and Remotely Sensed Vegetation Calibration Parameters**

367 Using the 11 years of destructive vegetation sampling from 3 fields near Mead, NE, we

368 found that the *GrWDRVI* was able to reasonably predict *SWB* when partitioning the data into

369 maize and soybean, irrigated and rainfed, and green-up/mature and senescence periods of crop

370 development (Figure 4 and Tables S2 and S3). Figure 4a and 4b illustrate the logistic functions

371 that were used to predict *SWB* for maize green-up (RMSE = 0.88 kg/m^2) and soybean green-up

372 (RMSE = 0.47 kg/m^2). We note that *SWB* relationships with *GrWDRVI* indicate that *GrWDRVI*

373 values less than 0.25 equated to the absence of *SWB*. During senescence, we found that a second

374 order power law function fit the data well. We found the maize senescence functions (DOY>

375 210) needed to be further partitioned by irrigated and rainfed conditions as limitations in soil

17

376   water will occur more quickly with mature plants that utilize the entire root zone. The resulting

377   functions for irrigated maize during senescence (RMSE = 0.75 kg/m^2) and rainfed maize during

378   senescence (RMSE = 0.92 kg/m^2) behaved well. For the soybean senescence function

379   (DOY>230), we found a single function behaved reasonably well for both irrigated and rainfed

380   conditions (RMSE = 0.45 kg/m^2). As expected from previous research (Ciganda et al, 2008;

381   Peng et al. 2011), we found that the *GrWDRVI* was a poor predictor of *SDB*/percent water

382   content of the vegetation. We will discuss the reasons and alternative strategies for estimating

383   *SDB* in section 4.2.

384   Using the derived relationships from the three study sites near Mead, NE, we applied the

385   equations to our two study sites near Waco, NE (~ 88 km from Mead, NE, Figure 5 and Tables 3

386   and 4). Figure 5 illustrates the time series of *SWB* using the 8 day MODIS product in

387   combination with the derived equations for both field sites. The figure also illustrates the

388   observed destructive sampling for 4 different sampling bouts. With the limited data, we found

389   the time series of *SWB* calculated from the MODIS data followed the expected green-up and

390   senescence *SWB* behavior for both the irrigated maize and soybean. The *GrWDRVI* derived *SWB*

391   largely captured the maximum observed value for both the irrigated maize (6.58 kg/m^2 vs. 6.2

392   kg/m^2) and irrigated soybean (2.61 kg/m^2 vs. 1.81 kg/m^2). The largest discrepancy was during

393   the maize green-up period (DOY 183) where the observed value was 2.4 kg/m^2 and ~4.0 kg/m^2

394   calculated from the *GrWDRVI*. While the derived equations behaved well for this limited

395   validation dataset, the equations should be tested at additional sites where other crop and soil

396   types may influence the function coefficients. Overall, the equations and regression fits resulting

397   in RMSE < 1 kg/m^2 are within the uncertainty of destructive biomass sampling in crops (Franz et

398   al., 2013; 2015). We note that 1 kg/m^2 is approximately equal to 1 mm of water or about 0.0033

399   cm^3/cm^3 of SWC over 300 mm. This indicates that for relatively small changes in *BWE* it will be

400   nearly indistinguishable from the noise in the CRNP measurements. By having general *SWB*

401   relationships (for eastern Nebraska) through time using the 8 day MODIS data, this could allow

402   for reasonable biomass corrections to N_0 with minimal effects (<0.01 cm^3/cm^3) on the overall

403   estimation of *SWC*.

404

3.3. Results of GSDE Soil Properties Error Propagation Analysis

406   In order to further assess the accuracy of our datasets, we synthetically altered the

407   parameters via a Monte Carlo error analysis. This was done using the GSDE soil parameters

408   (θ_{LW}, θ_{SOC}, and ρ_b) as compared to using local sampling (Figure 6). The analysis revealed that

409   for the given bounds of θ_{LW}, θ_{SOC}, and ρ_b, the maximum RSME was around 0.035 cm^3/cm^3 at a

410   *SWC* = 0.40 cm^3/cm^3. The asymmetric shape of all the curves is expected given the nonlinear

411   calibration function given in Eq. (4) and the bounded nature of soil moisture. We found that ρ_b

412   was by far the most sensitive parameter, followed by θ_{LW} and then θ_{SOC}. We expect the

413   influence of vegetation changes to be small on the overall accuracy of *SWC* (<0.01 cm^3/cm^3)

414   given the low RMSE described in section 3.2 (< 1 kg/m^2, which is ~1 mm of water or 0.0033

415   cm^3/cm^3 for a soil depth of 300 mm). We also note the critical factor in the error propagation

416   analysis is the assumed range of ρ_b, given that it is directly multiplied by the gravimetric water

417   content in the calibration function. Therefore, future sampling efforts or evaluations of available

418   datasets should seek to improve the accuracy of bulk density, meaning better estimates of the

419   mean, standard deviation, quantiles, and impact of land use practices on bulk density.

420

4. Discussion

4.1. Global Soil Calibration Parameters

The correlation between observed and GSDE clay content was very strong (Figure 2a) for all 61 sites in the CONUS except for the site in south central Texas (29.9492^o, -97.9966^o). The site occurred near a transition from vertisol to alfisol soil taxonomic groups; the site may have been improperly categorized (Table S1) or may have straddled a sharp gradient in clay contents. The strong correlation of the GSDE clay content with the observed values allowed us to use the GSDE clay content in understanding the correlation between clay content and lattice water organized by US soil taxonomic groups (Table 1). A strong correlation was only found for clay content and lattice water for the mollisol soil taxonomic group (see Greacen, 1981; Zreda et al., 2012). This strong correlation is significant because large areas of the Midwest and Great Plains regions of the United States are made up of mollisol soils. Globally, mollisol soils comprise about 7% of the land surface (United Nations 2007) but contain some of the highest productive grassland and crop areas (i.e. Central USA, Argentina, Central Eurasia). As such, the roving CRNP method remains applicable within grassland agricultural settings. No significant linear relationships with clay content were found for alfisol, aridisol, entisol, inceptisol, spodosol, or ultisol. Instead the mean value was assigned to the alfisol, aridisol, entisol, inceptisol, spodosol, and ultisol soil taxonomic groups when generating the CONUS map. We found the differences in most of the soil taxonomic mean values were statistically significant among different taxonomic groups given the small standard errors of the means (not shown but can be calculated from data in Table 1). The current analysis did not contain enough samples for the soil taxonomic groups of andisol, gelisol, histosol, oxisol, or vertisol to perform a linear regression or assign a mean value. We recommend future work to consider repeating the analysis for a larger dataset using

444   the FAO 2007 (United Nations 2007) soil classification of all 25 groups (also classified for our

445   sites in Table S1). Given the widespread interest in both the fixed and roving cosmic-ray

446   technology, a database of lattice water and clay content for each site could be developed. In

447   addition, warehouses like the Natural Resources Conservation Service (NRCS) in Lincoln, NE

448   contain stored samples from around the USA. This warehouse, and other soil repositories around

449   the globe, could be sampled in order to complete the global dataset for use by the cosmic-ray

450   community. Finally, the NRCS regularly updates the Soil Survey Geographic Database

451   (SSURGO), which contains higher spatial resolution and vertically resolved estimates of soil

452   texture and structure (i.e. clay content and bulk density). With the defined regression

453   relationships and soil taxonomic groups, better spatial maps of lattice water could be generated.

454   This may become important for applications of the rover at scales less than 1 km, such as using it

455   for applications in precision agriculture as well as increasing the reliability of the calibration

456   function.

457   The correlation between the observed and GSDE soil organic carbon was fairly poor,

458   particularly at the high end (> 4 wt. %). The history of land use is critical in determining carbon

459   pools and how they change through time (Post et al., 2000) and may not be well represented in

460   the GSDE. For arable lands, we note that organic carbon has a relatively small impact on the

461   calibration function as it is multiplied by several factors in the calibration equation, and is

462   relatively low and homogeneously distributed in the A-horizon due to land management

463   activities. However, in grassland and forest sites, high SOC amounts and strong SOC vertical

464   gradients typically exist in the top soil and may need to be quantified with local in-situ sampling

465   (e.g. Bogena et al., 2013). For rover survey experiments in these areas, we suggest that SOC be

466   sampled with composite samples, particularly between sites with varying land use histories

467   which can be identified using historical land cover maps.

468   Observed *in-situ* soil bulk density and GSDE bulk density exhibited a positive

469   relationship, albeit with low R^2. The poor fit and sensitivity of the parameter in the calibration

470   function increases the importance of identifying the range and variability of bulk density within

471   the rover sample domain. The variability shown here by the standard deviation of the bulk

472   density for the individual point samples within the 28 ha sample domain varied between 0.1 and

473   0.2 g/cm^3. Moreover, estimating the quantiles of bulk density at a site is key given the

474   propagation of error analysis presented in section 3.3. Thus, this result supports direct sampling

475   at key locations (along gradients of land use, soil taxonomic groups, etc.) to constrain the

476   quantiles of expected bulk density values. We also suggest that for rover surveys in the USA

477   (and regional elsewhere), additional higher resolution datasets like SSURGO, and its derivatives

478   (e.g. Polaris, Chaney et al., 2016), be used instead of the 1 km GSDE (in particular bulk density

479   data as a function of depth), as significant small scale variability may be averaged out. This may

480   be critical to account for in future roving CRNP research areas, such as precision agriculture or

481   small scale watershed monitoring where significant soil texture variation may exist at short

482   length scales. We note that this analysis is a first step in the incorporation of existing soil

483   databases that will no doubt continue to increase in spatial resolution and accuracy. Given the

484   increasing use of the roving CRNP technology, we anticipate similar analyses and procedures

485   will be undertaken on regional and local scales from existing and new databases as they become

486   available.

487

488   **4.2. Global Remotely Sensed Vegetation Calibration Parameters**

489   The comparison of 11 years of destructive vegetation samples from maize and soybeans

490   at 3 sites in eastern Nebraska indicated that the *GrWDRVI* was able to predict *SWB* in

491   agricultural fields, especially when partitioned into green-up vs. senescence and irrigated vs.

492   rainfed (Figure 4). However, as expected the *GrWDRVI* was unable to predict *SDB*. The main

493   reason is as the plants begin to dry out during the late summer and early fall, leaves lose their

494   chlorophyll and leaf structure begins to collapse thereby increasing reflected green and reducing

495   near-infrared light (Ciganda et al. 2008; Peng et al. 2011). This is exaggerated by a change in the

496   allocation of resources by the plant from leaves to grain, shifting where the majority of mass is

497   located and thus weakening the capacity for the *GrWDRVI* to predict *SDB*. This biological

498   investment of resources is more pronounced for maize than soybeans. As additional crops are

499   included in this analysis, the location and development of the fruit and seed will impact the

500   predictive relationships using vegetation indices. We refer to the reader to Duncan et al. (2015)

501   and Kumar et al. (2015) for a recent review of vegetation indices in remote sensing.

502   While the developed regression relationships for maize and soybean (Table S3) were

503   tested against independent biomass estimates from Waco, NE (Figure 5), we note that further

504   validation is needed. In terms of a strategy for estimating *SDB*, we suggest that proxies such as

505   crop type and growth stage be used. Franz et al. (2013 and 2015) found that in early stages,

506   maize and soybean had canopy water contents from 75-90 wt. %. By the end of senescence

507   before harvest, the canopy water contents were down to 25-35 wt. %, and thus very low *BWE*

508   and minimal impact on the low-energy neutron intensity. If growth stage is not directly known,

509   local meteorological observations, planting date, and crop variety can be used to compute

510   proxies (e.g. growing degree days) or simulated from crop models (Allen et al. 1998). We note

511   that having a reasonably accurate estimate of *SWB* and thus *BWE* (within ~ 1 kg/m^2) is all that is

512   required to have a relatively small impact (< 0.01 cm^3/cm^3) on the estimated *SWC*. Finally, we

513   note that this methodology is not applicable to areas with woody biomass. Following Franz et al.,

514   (2013), Hawdon et al., (2014), Baatz et al., (2015), and Coopersmith et al., (2014) we suggest

515   other vegetation relationships (i.e. *BWE* vs. N_0) be defined. However, given the relatively small

516   changes in *BWE* over the year in forests, we would expect small changes in N_0 through time. For

517   a more complete discussion of CRNP calibration in forests and estimates of time varying

518   changes in N_0 please see Bogena et al., 2013 and Heidbüchel et al., (2016).

519

520   **4.3. Roving CRNP Survey Recommendations**

521   With the continuing use of the roving CNRP we make the following recommendations on

522   best calibration and use:

523   1) Collect a series of full calibration datasets (θ_{LW}, θ_{SOC}, ρ_b, *SWB*, *SDB*) in different land

524   use areas and soil types in order to estimate the instrument specific slope and intercept for

525   dependence of N_0 with *BWE*.

526   2) In the rover sampling area, construct a map of land use including descriptions of:

527   vegetation/crop type, planting date, variety, rainfed vs. irrigated, and gravel vs. paved

528   roads vs. natural areas (see Chrisman and Zreda 2013 for a discussion of road influence

529   on neutron counts).

530   3) Collect a series of aggregate soil samples for soil organic carbon and lattice water around

531   the survey area. The samples should be collected across land use, soil texture, and soil

532   taxonomic groups. The GSDE or more local datasets like SSURGO and Polaris (Chaney

533 et al., 2016) in the USA can be used to select sites, cross validate samples, and fill in data

534 gaps.

535 4) Soil bulk density is the critical parameter in the calibration equations and overall

536 accuracy of the cosmic-ray neutron method. Bulk density should be collected locally

537 wherever possible to determine reasonable quantiles. More local datasets like SSURGO

538 and Polaris in the USA will likely perform better at smaller scales than the 1 km GSDE.

539 5) *SWC* validation datasets should be collected to independently assess the accuracy of the

540 rover survey results.

541

5. Summary and Conclusions

543 In this work, we developed a framework using globally available datasets for estimating

544 four $(\theta_{LW}, \theta_{SOC}, \rho_b, SWB)$ of the five key soil and vegetation parameters needed by the roving

545 cosmic-ray neutron method for estimating *SWC* in fast growing vegetation areas such as row

546 crop production in agricultural areas. The remaining crop vegetation parameter (*SDB*) can be

547 fairly well approximated by crop type, growth stage or simulated with crop models. The

548 accuracy of the GSDE soil database was tested against 61 calibration datasets from the CONUS.

549 We found that the 1 km GSDE compares well against observed clay content ($R^2 = 0.68$) but

550 much poorer against soil bulk density ($R^2 = 0.203$) and soil organic carbon ($R^2 = 0.175$).

551 Surprisingly, of the six soil taxonomic groups we investigated, only mollisols showed a

552 statistically significant correlation with clay content. The remaining five soil taxonomic groups

553 we investigated did show statistically different mean values. These mean values were used to

554 generate a map (not complete) of lattice water for the CONUS. From 11 years of destructive

555   sampling of maize and soybean fields in eastern Nebraska, we found that the 8-day 500 m

556   resolution MODIS derived *GrWDRVI* was highly correlated to *SWB*, particularly when

557   partitioning the fields into green-up vs. senescence and irrigated vs. rainfed (RMSE < 1 kg/m^2).

558   A propagation of error analysis indicated that the range of bulk density values was the most

559   sensitive calibration parameter. For the selected ranges, we found the GSDE vs. local sampling

560   resulted in a maximum RMSE of 0.035 cm^3/cm^3 at a *SWC* = 0.40 cm^3/cm^3. Finally, a list of best

561   practices for future roving CRNP experiments is provided.

562

573

574   **References**

575 Allen, R. G., L. S. Pereira, D. Raes, and M. Smith, Crop evapotranspiration. Guidelines for

576   Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56., Food and

577   Agriculture Organization of the United Nations, Rome, Italy. 1998.

578 Baatz, R., H. R. Bogena, H. J. H. Franssen, J. A. Huisman, W. Qu, C. Montzka, and H.

579   Vereecken, Calibration of a catchment scale cosmic-ray probe network: A comparison of

580   three parameterization methods, *Journal of Hydrology*, Vol. *516*, Pgs. 231-244.

581   doi:10.1016/j.jhydrol.2014.02.026, 2014.

582 Baatz R., Bogena H. R., Hendricks Franssen H. J., Huisman J. A., and Montzka C. An Empirical

583   Vegetation Correction for Soil Water Content Quantification Using Cosmic Ray

584   Probes, *Water Resources Research*, Vol. 51, No.4, Pgs. 2030-2046, 2015.

585 Beven K. J., and Cloke H. L. Comment on "Hyperresolution global land surface modeling:

586   Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et

587   al., *Water Resour. Res.*, Vol. 48, W01801, doi:10.1029/2011WR010982, 2012.

588 Binley A., Hubbard S. S., Huisman J. A., Revil A., Robinson D. A., Singha K., and Slater L. D.:

589   The Emergence of Hydrogeophysics for Improved Understanding of Subsurface

590   Processes Over Multiple Time Scales, *Water Resour. Res.*, Vol. 51,

591   doi:10.1002/2015WR017016, 2015.

592 Bogena H. R., Huisman J. A., Baatz R., Hendricks Frassen H. J., and Vereecken H.: Accuracy

593   of the Cosmic-Ray Soil Water Content Probe in Humid Forest Ecosystems: The Worst

594   Case Scenario, *Water Resour. Res.*, Vol. 49, Pgs. 5778-5791, doi:10.1002/wrcr.20463,

595   2013.

596 Bogena, H. R., J. A. Huisman, A. Guntner, C. Hubner, J. Kusche, F. Jonard, S. Vey, and H.

597   Vereecken (2015), Emerging methods for noninvasive sensing of soil moisture dynamics

598   from field to catchment scale: a review, *Wiley Interdisciplinary Reviews-Water*, Vol. *2*,

599   No. 6, Pgs. 635-647. doi:10.1002/wat2.1097, 2015.

600   Chaney, N. W., E. F. Wood, A. B. McBratney, J. W. Hempel, T. W. Nauman, C. W. Brungard,

601   and N. P. Odgers. POLARIS: A 30-meter probabilistic soil series map of the contiguous

602   United States. *Geoderma,* Vol. 274, Pgs. 54-67. doi:10.1016/j.geoderma.2016.03.025,

603   2016.

604   Chrisman B. B. and M. Zreda: Quantifying Mesoscale Soil Moisture with the Cosmic-Ray

605   Rover, *Hydrology and Earth System Sciences*, Vol. 17, No. 12, Pgs. 5097 – 5108, 2013.

606   Ciganda, V., Gitelson, A.A., and Schepers, J. Vertical Profile and Temporal Variation of

607   Chlorophyll in Maize Canopy: Quantitative "Crop Vigor" Indicator by Means of

608   Reflectance-Based Techniques, *Agronomy Journal*, Vol. 100, No. 5, Pgs. 1409-1417,

609   doi:10.2134/agronj2007.0322, 2008.

610   Coopersmith E. J., Cosh M. H., and Daughtry C. S. T.: Field-Scale Moisture Estimates using

611   COSMOS Sensors: A Validation Study with Temporary Networks and Leaf-Area

612   Indices, *Journal of Hydrology*, Vol. 519, Part A, Pgs. 637-643,

613   http://dx.doi.org/10.1016/j.jhydrol.20 14.07.060, 2014.

614   Crow, W. T., A. A. Berg, M. H. Cosh, A. Loew, B. P. Mohanty, R. Panciera, P. de Rosnay, D.

615   Ryu, and J. P. Walker: Upscaling Sparse Ground-Based Soil Moisture Observations For

616   The Validation Of Coarse-Resolution Satellite Soil Moisture Products. *Rev. Geophys*.

617   50. doi:10.1029/2011rg000372, 2012.

618   Dane, J. H., and C. G. Topp. Methods of Soil Analysis: Part 4 Physical Methods. Soil Science

619   Society of America. Madison, WI. 2002.

620   Desilets D., and Zreda M.: Footprint Diameter for a Cosmic-Ray Soil Moisture Probe: Theory

621   and Monte Carlo Simulations, *Water Resour. Res*., Vol. 49, Pgs. 3566–3575,

622   doi:10.1002/wrcr.20187, 2013.

623   Desilets D., Zreda M., Ferre T. P. A.: Nature's Neutron Probe: Land Surface Hydrology at an

624   Elusive Scale with Cosmic-Rays, *Water Resour. Res*., Vol. 56,

625   doi:10.1029/2009wr008726, 2010.

626   Dong J. N., Ochsner T. E., Zreda M., Cosh M. H., and Zou C. B.: Calibration and Validation of

627   the COSMOS Rover for Surface Soil Moisture Measurement, *Vadose Zone J*., Vol. 13,

628   No. 4, doi:10.2136/vzj2013.08.0148, 2014.

629   Duncan J.M.A, Dash J. and Atkinson P.M.: The potential of satellite-observed crop phenology to

630   enhance yield gap assessments in smallholder landscapes. Frontiers in Environmental

631   Science, Vol. 3, No. 56, doi: 10.3389/fenvs.2015.00056, 2015.

632   Entekhabi D., Njoku E. G., Neill P. E., Kellogg K. H., Crow W. T., Edelstein W. N., and

633   Kimball J.: The Soil Moisture Active Passive (SMAP) Mission, *Proceedings of the*

634   *IEEE*, Vol. 98, No. 5, Pgs. 704-716, 2010.

635   Franz T. E., Wang T., Avery W., Finkenbiner C., and Brocca L.: Combined Analysis of Soil

636   Moisture Measurements from Roving and Fixed Cosmic-Ray Neutron Probes for

637   Multiscale Real-Time Monitoring, *Geophys. Res. Lett*., Vol. 42, doi:10.1002/

638   2015GL063963, 2015.

639   Franz T. E., Zreda M., Ferre P. A., Rosolem R., Zweck C., Stillman S., Zeng X., and Shutt W. J.:

640   Measurement Depth of The Cosmic-Ray Soil Moisture Probe Affected by Hydrogen

641   From Various Sources. *Water Resour. Res*., Vol. 48, doi: 10.1029/2012WR011871,

642   2012.

643   Franz T. E., Zreda M., Rosolem R., Hornbuckle B. K., Irvin S. L., Adams H., Kolb T. E., Zweck

644   C., and Shuttleworth W. J.: Ecosystem-Scale Measurements of Biomass Water using

645   Cosmic-ray Neutrons, *Geophysical Research Letters*, Vol. 40, Pgs. 1–5,

646   doi:10.1002/grl.50791, 2013.

647   Gietelson, A.A., Wide Dynamic Range Vegetation Index for Remote Quantification of

648   Biophysical Characteristics of Vegetation, Journal of Plant Physiology, Vol. 161, Pgs.

649   165-173, 2004.

650   Greacen, E.L.: Soil water assessment by the neutron method, Australia: CSIRO. 1981.

651   Hawdon A., McJannet D., and Wallace J.: Calibration and Correction Procedures for Cosmic-

652   Ray Neutron Soil Moisture Probes Located Across Australia, *Water Resour. Res.*, Vol.

653   50, Pgs. 5029–5043, doi:10.1002/2013WR015138, 2014.

654   Heidbuchel, I., A. Guntner, and T. Blume. Use of cosmic-ray neutron sensors for soil moisture

655   monitoring in forests. *Hydrology and Earth System Sciences*, Vol. 20, No. 3, Pgs. 1269-

656   1288. doi:10.5194/hess-20-1269-2016. 2016.

657   Iwema, J., R. Rosolem, R. Baatz, T. Wagener, and H. Bogena. Investigating temporal field

658   sampling strategies for site-specific calibration of three soil moisture–neutron intensity

659   parameterisation methods. *Hydrology and Earth System Sciences,* Vol. 19, Pgs, 3203–

660   3216. doi:10.5194/hess-19-3203-2015. 2015.

661   Jenkins, J. C., D. C. Chojnacky, L. S. Heath, and R. A. Birdsey. National-scale biomass

662   estimators for United States tree species. Forest Science, Vol. 49, No. 1, Pgs. 12-35.

663   2003.

664 Kerr Y. H., Waldteufel P., Wigneron J. P., Delwart S., Cabot F. O., Boutin J., and Juglea S. E.:

665   The SMOS Mission: New Tool for Monitoring Key Elements of the Global Water Cycle,

666   *Proceedings of the IEEE*, Vol. 98, No. 5, Pgs. 666-687, 2010.

667 Köhli M., Schrön M., Zreda M., Schmidt U., Dietrich P., and Zacharias S.: Footprint

668   Characteristics Revised for Field-Scale Soil Moisture Monitoring with Cosmic-Ray

669   Neutrons, Water *Resources Research*. doi: 10.1002/2015WR017169, 2015.

670 Korres, W., T. G. Reichenau, P. Fiener, C. N. Koyama, H. R. Bogena, T. Comelissen, R. Baatz,

671   M. Herbst, B. Diekkruger, H. Vereecken, and K. Schneider, Spatio-temporal soil

672   moisture patterns - A meta-analysis using plot to catchment scale data, *Journal of*

673   *Hydrology*, Vol. *520*, Pgs. 326-341. doi:10.1016/j.jhydrol.2014.11.042, 2015.

674 Kumar, L., P. Sinha, S. Taylor, and A. F. Alqurashi. Review of the use of remote sensing for

675   biomass estimation to support renewable energy generation. Journal of Applied Remote

676   Sensing. Vol. 9, No. 28, doi:10.1117/1.jrs.9.097696, 2015.

677 Lv L., Franz T. E., Robinson D. A., and Jones S. B.: Measured and Modeled Soil Moisture

678   Compared with Cosmic-Ray Neutron Probe Estimates in a Mixed Forest, *Vadose Zone*

679   *Journal*, Vol. 13, No. 12, 2014.

680 McJannet D., Franz T. E., Hawdon A., Boadle D., Baker B., Almeida A., Silberstein R., Lambert

681   T., and Desilets D.: Field Testing of the Universal Calibration Function for

682   Determination of Soil Moisture With Cosmic-Ray Neutrons, *Water Resour. Res.,* Vol.

683   50, Pgs. 5235–5248, doi:10.1002/2014WR015513, 2014.

684   Mekonnen M. M. and Hoekstra A. Y.: The Green, Blue and Grey Water Footprint of Crops and

685   Derived Crop Products, *Hydrol. Earth Syst. Sci.*, Vol. 15, Pgs. 1577-1600,

686   doi:10.5194/hess-15-1577-2011, 2011.

687   Nguy-Robertson A., Gitelson A., Peng Y., Viña A., Arkebauer T., and Rundquist D.: Green

688   Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to

689   Achieve Maximal Sensitivity, *Agronomy Journal*, Vol. 104, No. 5,

690   doi:10.2134/agronj2012.0065, 2012.

691   Nguy-Robertson, A.L., and Gitelson, A.A.: Algorithms for estimating green leaf area index in C3

692   and C4 crops for MODIS, Landsat TM/ETM+, MERIS, Sentinel MSI/OLCI, and Venus

693   sensors, *Remote Sensing Letters*, Vol. 6, No. 5, doi: 10.1080/2150704X.2015.1034888,

694   2015.

695   Njoku E. G., and Entekhabi, D.: Passive Microwave Remote Sensing of Soil Moisture, *Journal of*

696   *Hydrology*, Vol. 184, No. 1, Pgs. 101-129, 1996.

697   Peichl, M., Leava, N. A. and Kiely, G. Above- and belowground ecosystem

698   biomass, carbon and nitrogen allocation in recently afforested grassland and adjacent

699   intensively managed grassland. Plant and Soil, Vol. 350, Pgs. 281-296. 2012.

700   Peng, Y., Gitelson, A.A., Keydan, G., Rundquist, D.C., and Moses, W. Remote Estimation of

701   Gross Primary Production in Maize and Support for a New Paradigm Based on Total

702   Crop Chlorophyll Content, *Remote Sensing of Envirionment*, Vol. 115, Pgs. 978-989,

703   doi:10.1016/j.rse.2010.12.001, 2011.

704 Post W. M., and Kwon K. C.: Soil Carbon Sequestration and Land-Use Change: Processes and

705   Potential, *Global Change Biology*, Vol. 6, Pgs. 317-327, doi:10.1046/j.1365-

706   2486.2000.00308.x, 2000.

707 Renzullo L. J., van Dijk A. I. J. M., Perraud J. M., Collins D., Henderson B., Jin H., Smith A. B.,

708   and McJannet D. L.: Continental Satellite Soil Moisture Data Assimilation Improves

709   Root-Zone Moisture Analysis for Water Resources Assessment, *Journal of Hydrology*,

710   Vol. 519, Part D, Pgs. 2747-2762, 2014.

711 Rosolem, R., W. J. Shuttleworth, M. Zreda, T. E. Franz, X. Zeng, and S. A. Kurc. The Effect of

712   Atmospheric Water Vapor on the Cosmic-ray Soil Moisture Signal. J. Hydrometeorol.

713   doi:10.1175/JHM-D-12-0120.1. 2013.

714 Rosolem R., Hoar T., Arellano A., Anderson J. L., Shuttleworth W. J., Zeng X., and Franz T. E.:

715   Translating Aboveground Cosmic-Ray Neutron Intensity to High-Frequency Soil

716   Moisture Profiles at Sub-Kilometer Scale, *Hydrol. Earth Syst. Sci.*, Vol. 18, Pgs. 4363-

717   4379, doi:10.5194/hess-18-4363-2014, 2014.

718 Shangguan W., Dai Y., Duan Q., Liu B., and Yuan H.: A global Soil Dataset for Earth System

719   Modeling, *J. Adv. Model Earth Syst.*, Vol. 6, Pgs. 249-263, doi:10.1002/2013MS000293,

720   2014.

721 Shuttleworth J., Rosolem R., Zreda M., and Franz T. E.: The Cosmic-Ray Soil Moisture

722   Interaction Code (COSMIC) for use in Data Assimilation, *Hydrol. Earth Syst. Sci.*, Vol.

723   17, Pgs. 3205-3217, doi:10.5194/hess-17-3205-2013, 2013.

724 Soil Survey Staff. Soil taxonomy: A basic system of soil classification for making and

725   interpreting soil surveys. 2nd edition. Natural Resources Conservation Service. U.S.

726   Department of Agriculture Handbook 436. 1999. Available online at

727   http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_051232.pdf.

728 Suyker A. E., Verma S. B., Burba G. G., and Arkebauer T. J.: Gross Primary Production and

729   Ecosystem Respiration of Irrigated Maize and Irrigated Soybean During a Growing

730   Season, *Agricultural Forest Meteorology*, Vol. 131, PGs. 180-190, 2005

731 United Nations Department of Economic and Social Affairs (United Nations). World Population

732   Prospects: The 2015 Revision.

733 United Nations Food and Agriculture Organization (FAO). How to Feed the World in 2050,

734   High *Level Expert Forum*, Rome, Italy, 2009.

735 United Nations Food and Agriculture Organization (United Nations). IUSS Working Group WRB,

736   World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources

737   Reports No. 103. FAO, Rome.

738 Vachaud G., Silans A. P. D. E., Balabanis P., Vauclin M.: Temporal Stability of Spatially

739   Measured Soil Water Probability Density Function, *Soil Science Society of America*

740   *Journal*, Vol. 49, Pgs. 822-828, 1985.

741 Viña A., Gitelson A. A., Nguy-Robertson A. L., and Peng Y.: Comparison of Different

742   Vegetation Indices for the Remote Assessment of Green Leaf Area Index of

743   Crops, *Remote Sensing of Environment*, Vol. 115, No. 12, Pgs. 3468-3478, 2011.

744 Wood E. F., Roundy J. K., Troy T. J., van Beek L. P. H., Bierkens M. F. P., Blyth E., de Roo A.,

745   Döll P., Ek M., Famiglietti J., Gochis D., van de Giesen N., Houser P., Jaffé P. R., Kollet

746   S., Lehner B., Lettenmaier D. P., Peters-Lidard C., Murugesu S., Sheffield J., Wade A.,

747   and Whitehead P.: Hyperresolution Global Land Surface Modeling: Meeting a Grand

748   Challenge for Monitoring Earth's Terrestrial Water, *Water Resources Research*., Vol. 47,

749   doi:10.1029/2010WR010090, 2011.

750   Zreda M., Desilets D., Ferre T. P. A., and Scott R. L.: Measuring Soil Moisture Content Non-

751   Invasively at Intermediate Spatial Scale Using Cosmic-Ray Neutrons, *Geophys. Res.*

752   *Lett.*, Vol. 35, No. 21, doi:10.1029/2008gl035655, 2008.

753   Zreda M., Shuttleworth W. J., Xeng X., Zweck C., Desilets D., Franz T. E., Rosolem R., and

754   Ferre P. A.: COSMOS: The Cosmic-Ray Soil Moisture Observing System. *Hydrol.*

755   *Earth Syst. Sci. Discuss*., Vol. 9, Pgs. 5405-4551, doi:10.5194/ hessd-9-4505-2012, 2012.

756

757   **Table Captions**

758   Table 1. Summary of mean, standard deviation of *in-situ* lattice water samples organized by USA

759   soil taxonomic groups. The table also summarizes a linear regression analysis using the GSDE

760   clay percent and *in-situ* sample. The last column indicates how the 1 km CONUS lattice water

761   map was generated. Note NA stands for not applicable because of a lack of data.

USA Soil Taxonomic Group	Mean Lattice Water (Wt. %)	Std. Lattice Water (Wt. %)	Number of Samples	Linear Regression Slope	Linear Regression Intercept	Linear Regression R^2	Linear Regression p value	GSDE Derived CONUS Lattice Water Product
Alfisol	4.31	1.36	9	6.09	-0.11	0.086	0.44330	Mean
Andisol	NA	NA	NA	NA	NA	NA	NA	NA
Aridisol	2.73	1.36	10	4.82	-0.15	0.095	0.38607	Mean
Entisol	1.47	0.93	5	2.48	-0.14	0.233	0.41064	Mean
Gelisol	NA	NA	NA	NA	NA	NA	NA	NA
Histosol	NA	NA	NA	NA	NA	NA	NA	NA
Inceptisol	4.98	0.28	2	NA	NA	NA	NA	Mean
Mollisol	3.18	1.22	24	1.03	0.11	0.539	0.00004	Linear
Oxisol	NA	NA	NA	NA	NA	NA	NA	NA

Spodosol	2.68	2.10	4	3.45	-0.11	0.020	0.85919	Mean
Ultisol	2.82	2.33	6	0.28	0.20	0.229	0.33672	Mean
Vertisol	5.18	NA	1	NA	NA	NA	NA	NA
ALL	3.16	1.58	61	1.68	0.09	0.183	0.00066	NA

762

763

Table 2. Top) Summary of mean difference between *in-situ* samples and GSDE values (Figure 3) for bulk density, lattice water and organic carbon. Bottom) Summary of covariance matrix of difference between *in-situ* values and GSDE values. The mean difference and covariance data were used in an error propagation analysis illustrated in Figure 6.

	Bulk Density (g/cm^3)	Lattice Water (Wt. %)	Organic Carbon (Wt. %)
Mean Difference of in-situ value - GSDE value	-0.10035	-0.05789	-0.07077
Covariance matrix of in-situ value - GSDE value			
	Bulk Density (g/cm^3)	Lattice Water (Wt. %)	Organic Carbon (Wt. %)
Bulk Density (g/cm^3)	0.0386	-0.0567	-0.2077
Lattice Water (Wt. %)		1.6745	0.3624
Organic Carbon (Wt. %)			3.5810

768

769

770

771

772

773

774

775

776

777

778   Table 3. Summary of 2014 *GrWDRVI* and calculated standing wet biomass for irrigated maize

779   and irrigated soybean fields near Waco, NE. Note that the senescence equation was applied to

780   DOY 209 for the irrigated maize field as planting date and development can vary locally. The

781   drop in *GrWDRVI* between DOY 201 and 209 is a clear indicator of change in plant growth stage

782   that can be used on a field by field basis.

DOY (2014)	GrWDRVI, Irrigated-Maize	GrWDRVI-Irrigated Soybean	Calculated Standing Wet Biomass-Irrigated Maize (kg/m^2)	Calculated Standing Wet Biomass-Irrigated Soybean (kg/m^2)
153	0.23	0.23	0.00	0.00
161	0.24	0.24	0.00	0.00
169	0.32	0.28	0.53	0.06
177	0.57	0.54	4.69	1.25
185	0.55	NA	4.33	NA
193	0.63	0.63	5.63	1.91
201	0.61	0.71	5.34	2.48
209	0.55	0.73	6.50*	2.61
217	0.57	0.74	6.58	2.67
225	0.50	0.73	6.27	2.61
233	0.47	0.74	6.07	NA
241	0.40	0.68	5.38	2.89
249	0.43	0.64	5.73	6.77
257	0.27	0.47	1.44	6.07
265	0.25	0.44	0.00	5.83
281	0.21	0.28	0.00	2.02
289	0.21	0.26	0.00	0.78
297	0.20	0.25	0.00	0.00

783

784

785

Table 4. Summary of 2014 observed standing wet biomass for irrigated maize and irrigated soybean fields near Waco, NE. The observations represent the aggregation of 18 plants collected at 6 different locations across the field on the sampling date.

DOY (2014), Irrigated Soybean	Observed Standing Wet Biomass- Irrigated Soybean (kg/m^2)	DOY (2014), Irrigated Maize	Observed Standing Wet Biomass- Irrigated Maize (kg/m^2)
167	0.19	161	0.13
196	1.63	183	2.40
211	1.81	217	6.22
259	1.63	259	0.30

789

790

791

792

793

794

795

796

797

798

799

800

801   Table S1. Summary of *in-situ* and GDSE soil information for 61 CONUS study sites (see

802   supplemental material zip file).

803

804   Table S2. Summary of observed standing wet biomass and MODIS derived *GrWDRVI* for each

805   of the 3 fields near Mead, NE (see supplemental material zip file).

806

807   Table S3. Summary of derived equations estimating standing wet biomass from *GrWDRVI* for

808   maize and soybean partitioned into irrigated and rainfed areas and green-up (DOY< 210 for

809   maize, DOY<230 for soybean) and senescence. Destructive biomass data is aggregated from 3

810   fields near Mead, NE between 2003-2013 (Table S2). We note that the maize and soybean

811   functions were bounded to provide realistic behavior at the observed *GrWDRVI* and destructive

812   vegetation sampling bounds. See main text for details.

813

814

815

816

817

818

819

820

Figure Captions

Figure 1. Map of soil taxonomic classification map over the Continental United States of America using the twelve USA soil taxonomic orders (data source FAO 2007 and personal communication with M. Kuzila). Note gelisols are not present in the CONUS. Black dots indicate 61 locations where we have *in-situ* composite/average samples for soil bulk density, soil lattice water, soil organic carbon, and clay weight fraction collected over a 12.6 ha circle and averaged over the top 30 cm (Table S1).

Figure 2. Comparison between 61 *in-situ* composite sample and GSDE value from the closest pixel for a) clay weight percent b) soil bulk density, and c) soil organic carbon. d) Comparison between *in-situ* lattice water and derived values using GSDE clay weight fraction and soil taxonomic orders. See Table 1 for summary of data by taxonomic group, Table S1 for raw data, and Table 2 for statistical summary of differences between *in-situ* and GSDE product. Note error bars denote +/- 1 standard deviation.

Figure 3. Derived 1 km resolution lattice water weight percent map using the GSDE clay percent and regression analyses organized by soil taxonomic classification. See Table 1 for estimates of the mean, standard deviation, and linear regression vs. clay percent organized by taxonomic

839   group. Black dots indicate 61 locations where we have *in-situ* composite/average samples for soil

840   bulk density, soil lattice water, soil organic carbon, and clay weight fraction collected over a 12.6

841   ha circle and averaged over the top 30 cm (Table S1). Missing areas indicate surface water

842   bodies or soil taxonomic groups with no or limited *in-situ* lattice water sampling (see Table 1).

843

844   Figure 4. Relationship between *GrWDRVI* and observed standing weight biomass for maize (a,

845   c) and soybean (b, d) partitioned into green-up (DOY< 210 for maize, DOY<230 for soybean)

846   and senescence. Destructive vegetation data is aggregated from 3 fields near Mead, NE between

847   2003-2013 (Table S2). The regression coefficients and equations are summarized in Table S3.

848   Note that the maize and soybean functions were subject to the constraints in order to provide

849   realistic behavior at the observed *GrWDRVI* and destructive vegetation sampling bounds. See

850   main text for details.

851

852   Figure 5. Time series of standing wet biomass for two study sites (irrigated maize and irrigated

853   soybean) near Waco, NE over the 2014 growing season. The graph contains the observed *in-situ*

854   sampling in addition to the *GrWDRVI* estimates using the equations summarized in Table S3.

855   See Table 3 for *GrWDRVI* values and Table 4 for *in-situ* estimates.

856

857   Figure 6. Propagation of error analysis using Monte Carlo simulations of 100,000 soil parameter

858   datasets of true soil parameters (i.e. soil bulk density, lattice water, soil organic carbon) and

859   perturbed parameters with matching mean differences and covariance matrix between *in-situ*

860   samples and GSDE derived parameters (see Table 2). Three error metrics are presented across a

861   range of neutron counts (and thus *SWC* values). Note that soil bulk density was constrained to

862   1.2-1.5 g/cm^3, lattice water was constrained from 1-8 wt. %, soil organic carbon was constrained

863   from 0-8 wt. %, and soil water content was constrained from 0.03-0.45 cm^3/cm^3. Simulated and

864   calculated values outside of these bounds were either reset to the minimum or maximum or

865   removed from the Monte Carlo statistics. A minimum threshold of 70% of simulated cases were

866   used to compute error statistics.