Supplementary 2

Further details on optically stimulated luminescence dating

2.1 Estimation of the error of the water content

The error of the water content of 0.06 (1 sigma, ratio of wet sample weight to dry sample weight; section 3.2.6/main text) was derived as follows: The texture of the OSL samples was silt loam to loam (cf. Fig. S-1.3 of supplementary 1). Blume et al. (2013) compiled total pore volume data for sands, silts, clays and loams as well as pore volume data differentiated into coarse, medium and fine pores for sands, silts and clays. While total pore volumes for loams are 43 ± 13 %, they are 47 ± 9 % for silts. For edaphically drained sites in the humid zones, where water uptake by plants is no issue, it is reasonable to assume that fine pores (ø < 10 µm) are always filled with water, whereas coarse pores are filled only during rainfall-infiltration events, which are negligible over the whole dating period. Medium pores would be the pore volume that may or may not be filled by water. Here we take the data for silt of Blume et al. (2013) as the best approximation to the Datthausen samples. They comprise 15 ± 10 % coarse pores, 15 ± 7 % medium pores and 15 ± 5 % fine pores, i.e. ca. one third to each pore class. Assuming 2.65 g/cm³ for the mineral component, 0 g/cm² for air filled pores and 1 g/cm³ for water filled pores the ratio of a sample's wet weight to dry weight (Δ) may be modelled. For the whole spectrum of possible total pore volumes the mean Δ-values range between 1.12 and 1.25 (cf. supplementary 3). For a silt sample of 47 % total pore volume with ca. one third to each of the three pore classes a minimum Δ of 1.12 may be assumed, if all fine pores are filled with water, and a maximum Δ of 1.22 may be assumed, if also all medium pores are filled with water. In not water-logged sites, i.e. where coarse pores are not water-filled, this would cover the total possible range (3 sigma). As most sampled horizons exhibit massive to fine platy structure and are moreover overlain by 2 – 3 m loess, the pore volume is likely below average. Some samples, however, show characteristics of iron- and manganese dynamics (cf. section 4.1/main text) which are typical for stagnant conditions and which likely prevailed seasonally during Pleniglacial times, e.g. when the upper part of the sediment-soil profile thawed in the summer time in stadial periods, or in humid times of the year in interstadial periods. Assuming, for a sample with 47 % pore volume, 20 % of the whole dating period as water-logged (80 % Δ 1.20, 20 % Δ 1.33) would result in 1.23 for the whole dating period. Summarizing, considering an error of 0.06 on the 1 sigma error level, on

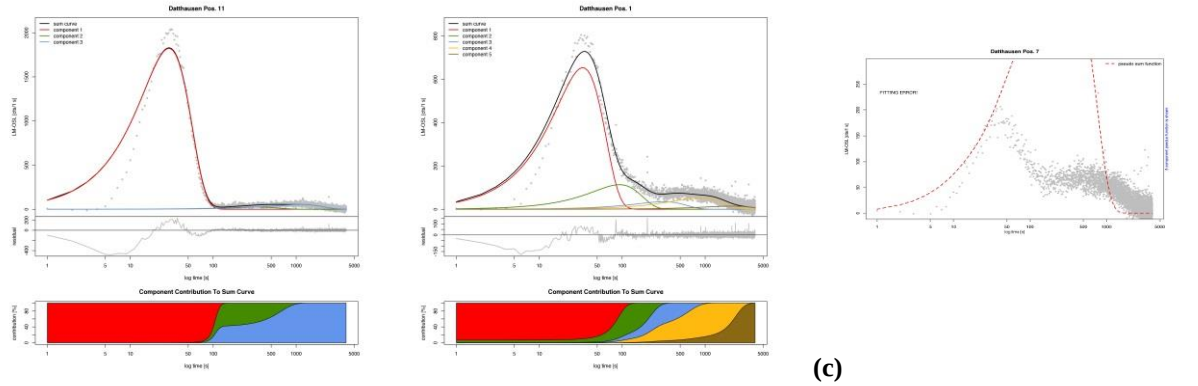
which the OSL-ages are conventionally quoted, appears reasonable for the samples investigated here.

2.2 Pretests to optimize the SAR-protocol for BLSL dating

To optimize the SAR-protocol as summarized in Table 3/main text the Datthausen samples were subjected to different pretests as described in the following sections.

2.2.1 Linearly modulated (LM) luminescence curves

Quartz samples dominated by medium and slow components of the BLSL curve may underestimate the true palaeodose (e.g. Li and Li, 2006). In order to test whether the Datthausen samples possess a dominant fast component beneficial for BLSL dating, we performed linearly modulated (LM) luminescence measurements on six natural aliquots, preheated for 10 s at 220 °C and 240 °C (3 aliquots each). The ramping time was set to 4000 s using a resolution of one channel per second. To also determine possible IRSL contribution, the aliquots were twice given a laboratory beta dose of 1500 s (ca. 140 Gy), read out using once BLSL at 125 °C and once IRSL at 125 °C. The IRSL contribution was determined by comparing the gross signal counts. Component deconvolution was done using the R package 'Luminescence' (Kreutzer et al., 2012).



(a)	**(b)**	**(c)**
position = #11	position = #1	position = #7
preheat-1 = 240 °C	preheat-1 = 220 °C	preheat-1 = 240 °C
IRSL-125 °C * 100/BLSL @ 125 °C = 0 %	IRSL-125 °C * 100/BLSL @ 125 °C = 3.6 %	IRSL-125 °C * 100/BLSL @ 125 °C = 24.6 %

Fig. S-2.1. BLSL-LM measurements of three different aliquots of sample HDS-1663. **(a)** Aliquots with pure quartz separates exhibit a dominant fast component. **(b)** Aliquots with a measurable IRSL-signal (here 3.6 %; determined by an IRSL test following the LM measurement; gross signal counts, no background subtraction; integral 0 – 0.16 s) show also further components. **(c)** Aliquots with significant IRSL contamination do not allow for fitting a dominant fast component.

The measurements showed that the Datthausen quartz possesses a dominant fast component. But increasing IRSL contents, resulting presumably from feldspar impurities, correlate with an increase of additional components in the BLSL curve (Fig. S-2.1). As a consequence, for SAR-measurements we eliminated all aliquots with IRSL proportion ≥ 10 % (cf. section 3.2.5/main text). Additionally, we used an early and short integral for D_e-determination (0 – 0.16 s, first data channel) and an early integral for background subtraction (1.12 – 1.28 s, channels 5 to 6) to avoid any malign effects of potential IRSL contribution to the luminescence signal, which increases relatively in size with increasing readout time.

2.2.2 Dose recovery tests

Prior to dating, we performed dose recovery tests on artificially bleached and dosed samples. The tests were of two different types: (a) combined dose-recovery and preheat tests and (b) combined dose-recovery, preheat and normalization-dose tests. The SAR protocol was as described in section 3.2.4/main text. We tried several intervals for D_e-determination and late light subtraction (cf. Aitken and Xie, 1992), among others an early integral 0 – 0.16 s (channel 1 – 1) with an early background for late-light subtraction of 0.8 – 0.96 s (channel 5 – 6), as well as a longer D_e-interval 0 – 0.64 s (channel 1 – 4) with a slightly later background for late-light subtraction of 1.12 – 1.28 s (channel 7 – 8). As results were in accordance within error margins, we present only the results of the early D_e-interval with the early late-light subtraction, which was also used for the dating samples. For natural samples analysis of an early D_e-interval is beneficial as it represents the easiest to bleach component of the OSL-shine down.

Dose recovery tests to optimize the preheat temperature

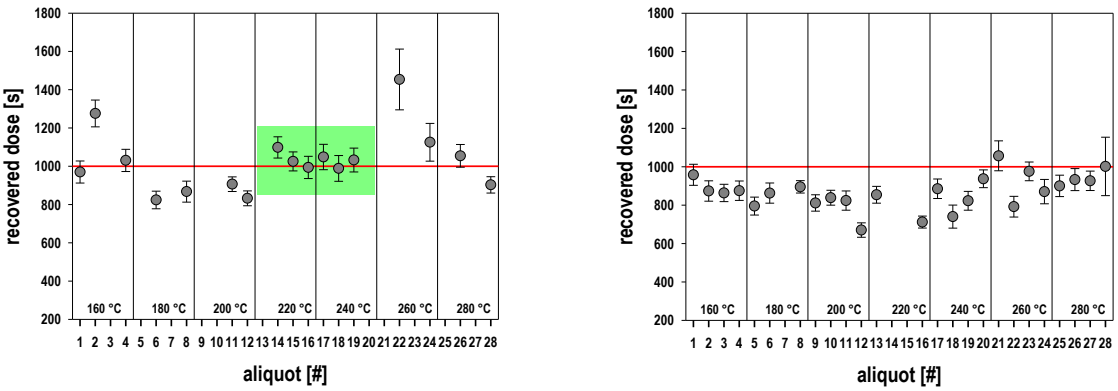
We performed two different combined dose-recovery and preheat tests (test-1 and test-2) to optimize the treatment of samples at elevated temperatures prior to BLSL readout. Prior to each test, samples were bleached for 3 hours using a Dr. Hönle solar simulator with samples placed on a water-cooled copper-plate (ca. 18 °C) and covered by a UVILEX glass filter to align the lamp's spectrum to the natural solar spectrum. Afterwards samples were stored in the dark for at least one day to allow for potential electron flux to decay. In the luminescence reader, samples received a laboratory dose of ca. 95 Gy (1000 s beta irradiation), as this order of magnitude was assumed to approximate the natural dose of the Datthausen samples. Regeneration dose points were at 20 %, 40 %, 70 %, 100 %, 130 %, 160 %, 20 %, and 0 %.

Finally, a regeneration dose of 1000 s (test-1) or 700 s (test-2) was administered a second time followed by an IRSL-readout at 125 °C preceding the usual BLSL-readout at 125 °C (IRSL test). The normalization dose was equal to the laboratory dose of 1000 s.

In test-1, preheat-1 was varied from 160 °C to 280 °C in steps of 20 °C (4 aliquots each), while preheat-2 was kept constant at 160 °C. In test-2, preheat-2 was varied simultaneously with preheat-1, from 160 °C to 280 °C (4 aliquots each). The tests were performed on 28 aliquots each. The preheat tests showed that increasing preheat-2 contemporaneous with preheat-1 tends to slightly underestimate the administered laboratory doses (Fig. S-2.2b). Keeping preheat-2 low at 160 °C, a preheat-1 of 220 °C or 240 °C delivered the best results (Table S-2.1; Fig. S-2.2a).

Table S-2.1: Group results of the combined dose-recovery and preheat test-1 (preheat-2 = 160 °C) for preheat-1 of 220 °C and 240 °C. Numbers in black: D_e-integral 0-0.16 s, late light 0.8-0.96 s. Numbers in blue: D_e-integral 0-0.64 s, late light 1.12-1.28 s.

	preheat-1 = 220 °C	preheat-1 = 240 °C
mean	1039.2 ± 53.70 s	1023.2 ± 30.83 s
	985.3 ± 55.47 s	1027.0 ± 39.67 s
weighted mean	1039.2 ± 51.71 s	1023.9 ± 30.14 s
	983.4 ± 52.73 s	1026.2 ± 39.63 s
central value	1041.8 ± 31.25 s	1024.8 ± 37.74 s
	986.2 ± 28.79 s	1027.7 ± 36.81 s

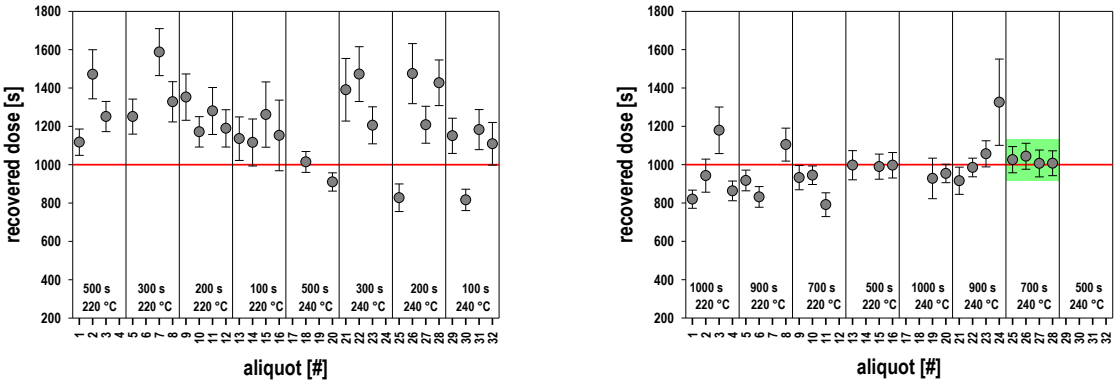


(a) test-1 **(b) test 2**

Fig. S-2.2. Combined dose recovery and preheat tests on 28 aliquots of sample HDS-1663 with laboratory and normalization dose of 1000 s and varying preheat-1 (160 – 280 °C). **(a)** Test-1 with constant preheat-2 of 160 °C. **(b)** Test-2 with preheat-2 varied like preheat-1. Results for D_e-interval 0 – 0.16 s (data channel 1 – 1) with late-light subtraction of 0.8 – 0.96 s (channel 5 – 6). Red line denoting expected value. Expected value of ca. 95 Gy (1000 s beta irradiation) met best for preheat-1 of 220 °C or 240 °C with preheat-2 of 160 °C (test-1, green area). Missing aliquots (here e.g. #13 and #20 in test-1) not meeting the threshold criteria, as detailed in section 3.2.5/main text.

Dose recovery tests to optimize the normalization dose

In a further series of two tests we combined a dose-recovery test with a normalization-dose test (test-3 and test-4) to check whether it would be beneficial to reduce the measurement impact on the aliquots by decreasing the size of the normalization dose. In both tests we applied a preheat-2 of 160 °C together with a preheat-1 of 220 °C as well as 240 °C (16 aliquots each). In test-3 we varied the normalization dose: 500 s, 300 s, 200 s and 100 s (4 aliquots each); in test-4 we varied the normalization dose: 1000 s, 900 s, 700 s and 500 s. The tests showed that low normalization doses overestimate the administered laboratory dose and result in larger errors. The latter is owed to the fact that the Datthausen quartz shows only moderately bright BLSL signals (Fig. S-2.3a). Higher normalization doses of ≥ 700 s delivered the best results (Fig. S-2.3b). For dating we chose a preheat-1 of 240 °C for 10 s, a preheat-2 of 160 °C for 10 s and a normalization dose of 700 s (ca. 65 Gy), which takes less measurement time than 1000 s.



(a) test-3 **(b) test-4**

Fig. S-2.3. Combined dose-recovery, preheat and normalization-dose tests on 32 aliquots of sample HDS-1663 with preheat-1 of 220 °C and 240 °C (16 aliquots each) and preheat-2 of 160 °C. **(a)** Test-3 with normalization doses of 500 s, 300 s, 200 s and 100 s beta-irradiation time. **(b)** Test-4 with normalization doses of 1000 s, 900 s, 700 s and 500 s. Data analysis and graphical display like Fig. S-2.2. Preheat-1 of 160 °C, preheat-2 of 240 °C and a normalization dose of 700 s (test-4, green area; less measurement time than 1000 s) used for SAR-protocol of the dating samples.

2.2.3 Test of BLSL-signal stability

In view of an unavoidable minor IRSL-signal contribution to the BLSL signal we performed tests to scrutinize whether the luminescence signal used for dating is stable, at least on laboratory time scales. The test was motivated by the observation that feldspar, when measured with UV-transmissive detection filters may lead to dose and therefore OSL-age underestimation (e.g. Wallinga et al., 2000; Timar et al., 2010) due to anomalous signal fading (Wintle, 1973) of the feldspar component. For the signal-stability test we used a sequence of SAR sub-cycles basically following the protocol of Table 3/main text, but in each SAR sub-cycle the sample was administered an identical laboratory dose of 700 s (ca. 67 Gy) starting with the first SAR sub-cycle. In between laboratory irradiation and preheat-1 pauses of variable lengths were intercalated in the different SAR sub-cycles which delayed the BLSL readout from 0 s (immediate readout) to 60 s, 600 s, 1200 s, 1800 s, 3600 s, 7200 s and 28800 s (readout delayed for 8 h) as illustrated in Fig. S-2.4. To account for possible statistical scattering of the data, immediate readout (0 s delay) occurred 3 times both at the beginning (sub-cycles 1 to 3) and at the end of a complete measurement (sub-cycles 12 to 14). Prior to the signal-stability test we determined the relative contribution of the IRSL signal to the BLSL signal for 10 aliquots of sample HDS-1663.

For the signal-stability test we chose 5 aliquots below the accepted threshold value of 10 % IRSL-signal contribution (see section 3.2.5/main text) which had a sufficiently strong BLSL signal of several 10^3 counts in the early D_e-integral 0 – 0.16 s. The IRSL-signal contribution was between 0.3 % and 5.2 % (cf. Fig. S-2.4e and c). Data analysis is based on the L_x/T_X-values as delivered by the software package Analyst (Duller 2013), which for each of the 14 sub-cycles represent the strength of the BLSL-signal for the repeatedly administered laboratory dose of ca. 67 Gy, divided by the signal of the corresponding normalization dose (test dose).

Fig. S-2.4 illustrates the L_x/T_x-values (upper part of each graph) as well as the L_x/T_x-values as normalized to the first L_x/T_x-value of a measurement (lower part of each graph). Fig. S-2.4a and b show two aliquots which during the course of the SAR measurement tend to deviate from 1, i.e. the SAR-protocol does not properly correct the sensitivity changes. It is likely that due to the poor recycling behavior one or both aliquots, had they undergone a SAR-dating measurement, would have been deleted from the data set. Apart from that, none of the here investigated aliquots shows any dependency of the SAR-corrected BLSL signal on the delay time. Therefore, the BLSL signal was regarded as suitable for dating.

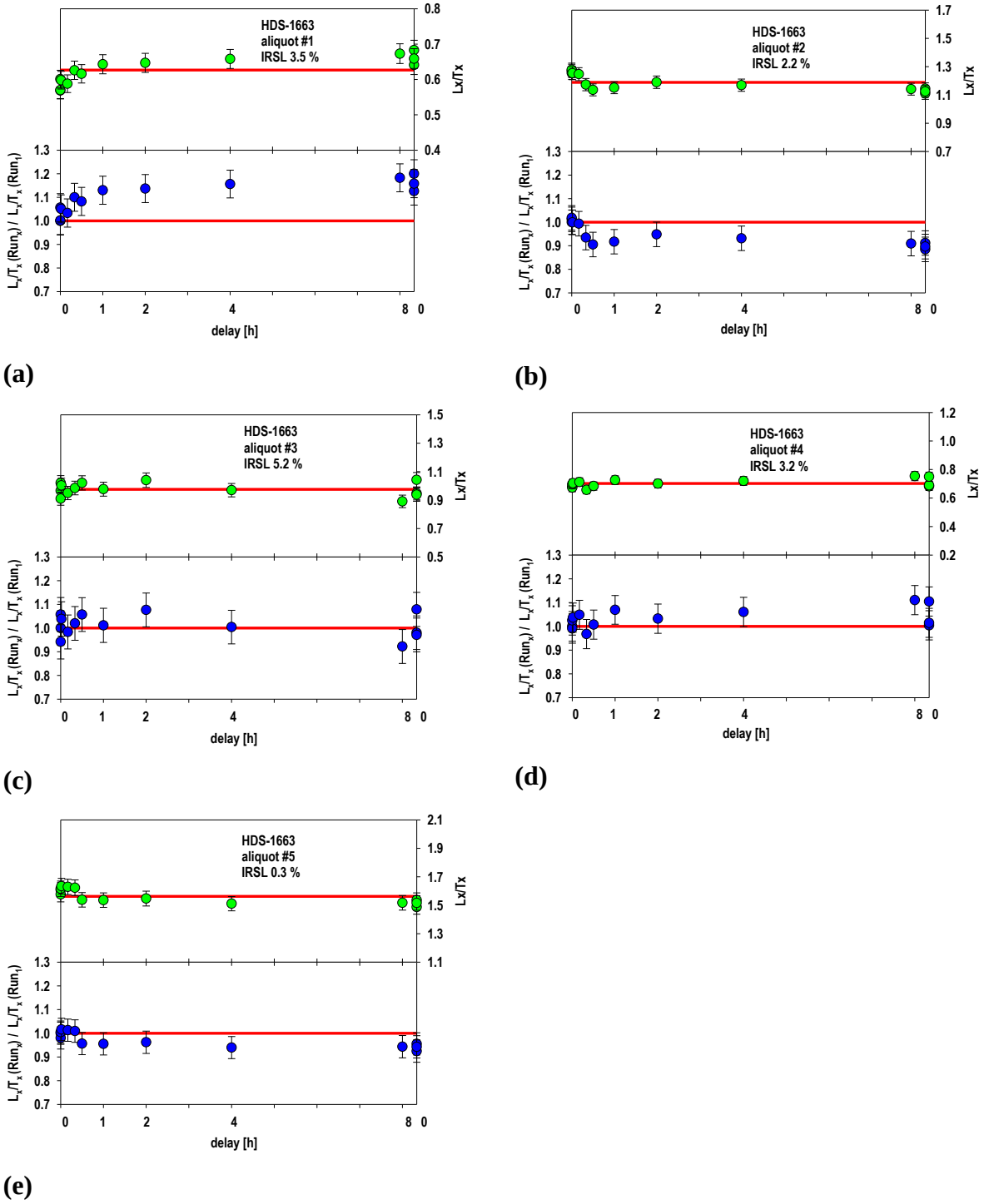


Fig. S-2.4. Test of BLSL-signal stability on laboratory time scales. Aliquots with IRSL-signal contribution to the BLSL-signal (integral 0 – 0.16 s) varying between 0.3 % and 5.2 %. Preheat-1 and BLSL readout delayed up to 8 h after laboratory irradiation of 700 s (ca. 67 Gy). Pauses being 0 s (3 times), 60 s, 600 s, 1200 s, 1800 s, 3600 s, 7200 s, 14400 s, 28800 s and 0 s (3 times). Upper part of each graph showing L_x/T_x-curves (green circles with mean value as red reference line). Lower part of each graph showing L_x/T_x-values normalized to L_x/T_x-value of first SAR subcycle (blue circles with red line indicating unity).

2.3 Follow-up measurements and analyses

2.3.1 Dose-recovery tests to evaluate sigma b

A reviewer encouraged us to perform an additional dose-recovery test on 24 – 36 aliquots in order to confirm that the samples' overdispersion does not exceed 20 % which was assumed as an appropriate parameter value (cf. sigma b of 0.2) by us for the minimum-age model approach. Basically, the follow-up dose-recovery tests follow the design outlined in section 2.2.2/supplementary (i.e. bleaching under a solar lamp Dr. Hönle SOL 2 and laboratory irradiation within the luminescence reader), but this time we confined to the preheat procedure as used for the dating samples, i.e. with preheat-1 240 °C and preheat-2 160 °C, 10 s each. Apart from the artificial dosing, the measurement sequence conforms to the measurement sequence as used for dating. We carried out altogether four of these follow-up dose-recovery tests: (1) sample HDS-1663, LAB 1000 s (ca. 89.5 Gy) 32 aliquots; (2) samples HDS-1660 to HDS-1664, each sample with a beta laboratory dose (LAB) corresponding in size to the sample's natural dose, 8 aliquots each sample; (3) HDS-1663, LAB 680 s (ca. 60.8 Gy), corresponding to the samples natural dose; (4) HDS-1663, like (3) but using aluminium cups of an earlier produced batch (cf. section 2.3.2/supplementary). Data analyses occurred with the software program 'Luminescence Analyst', version 4.31.7 (Duller 2015). In each case the overdispersion was below 10 %, suggesting that sigma b of 0.11 (Cunningham and Wallinga 2012) or 0.2 (Arnold and Roberts 2009) are likely reasonable assumptions. In this context it should be noted that in a natural environment sigma b of a well bleached sample will most likely exceed the value of overdispersion as gained from a dose-recovery test of an artificially bleached and dosed sample, as other factors, such as less perfect bleaching or microdosimetric heterogeneiity or others will add to the value gained in a laboratory environment.

However, apart from that confirming result, the follow-up dose-recovery tests in spring 2016 were not in perfect agreement with the dose-recovery tests carried out during the pretests in 2013. Although most aliquots met the expected value within an error margin of 10 %, the follow-up dose-recovery tests generally tended to underestimate the administered laboratory doses slightly. The dose-recovery tests in 2013 had shown that the Datthausen samples are temperature-sensitive (preheat procedures) with respect to the recovered dose (cf. 2.2.2/supplementary). Further, it is known that different luminescence readers possess slightly different heating efficiency at the sample position (Kadereit and Kreutzer 2013). It

seems that the heating efficiency of the here used reader (No. 245) may have changed between 2013 and 2016 for some unknown reason, thus rendering the original dating protocol now less appropriate. As a consequence of the unexpected results of the follow-up dose-recovery tests we checked the heating thoroughly, with the kind help of Per Sørensen/Risø-DTU. Judging by the results of the electronic test measurements, the reader works presently correct, like the other two readers in the laboratory, despite the fact that the positions of the TL-peaks of a sample vary slightly between readers, and may vary slightly on one reader over time. Further, we checked the TL-curves recorded during the preheat-2 for the pretest, dating and follow-up measurements. As the sequences were not run one aliquot at a time and laboratory irradiation of all (usually 32) aliquots lasted for more than 6 hours this comparison could be performed only for the last aliquots on the turntable of which the laboratory induced TL-peaks had not yet faded by the time when the BLSL-readout was performed. However, we could not observe any severe differences in the peak position of the 110 °C quartz peak (here mostly 94 °C, partly 101 °C; the difference corresponding to one data channel of 1.28 s/channel in the range 0 – 165 °C, heating rate 5 K/s) which could explain the slight differences between the earlier and the later dose-recovery tests.

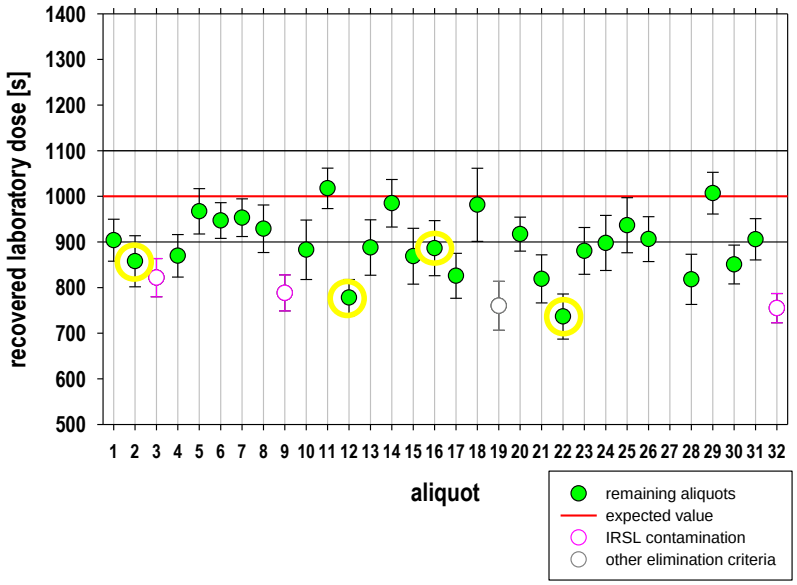


Fig. S-2.5. Dose-recovery test with 32 aliquots of sample HDS-1663 with an administered laboratory dose (LAB) of 1000 s (ca. 89.5 Gy). Although most aliquots meet the expected value (cf. red line) within an error margin of 10 % (cf. black lines), LAB is slightly underestimated by ca. 10 % (N = 27; mean with standard deviation and standard error 896.9 ± 67.5 ± 13.0 s; central-age D_e 900.9 ± 13.2 s; overdispersion = 4.9 ± 0.3 %). Yellow circles marking remaining aliquots (green), which do not comply with the pIR-BLSL recycling ratio threshold of 0.9 (alternative / additional test of scrutinizing IRSL contamination; cf. section 2.3.3/supplementary). Mean value underestimating expected value by ca. 8 % (N = 23; mean = 911.2 ± 57.3 ± 11.9 s; central-age D_e = 917.7 ± 12.0 s; overdispersion = 2.9 ± 0.4 %).

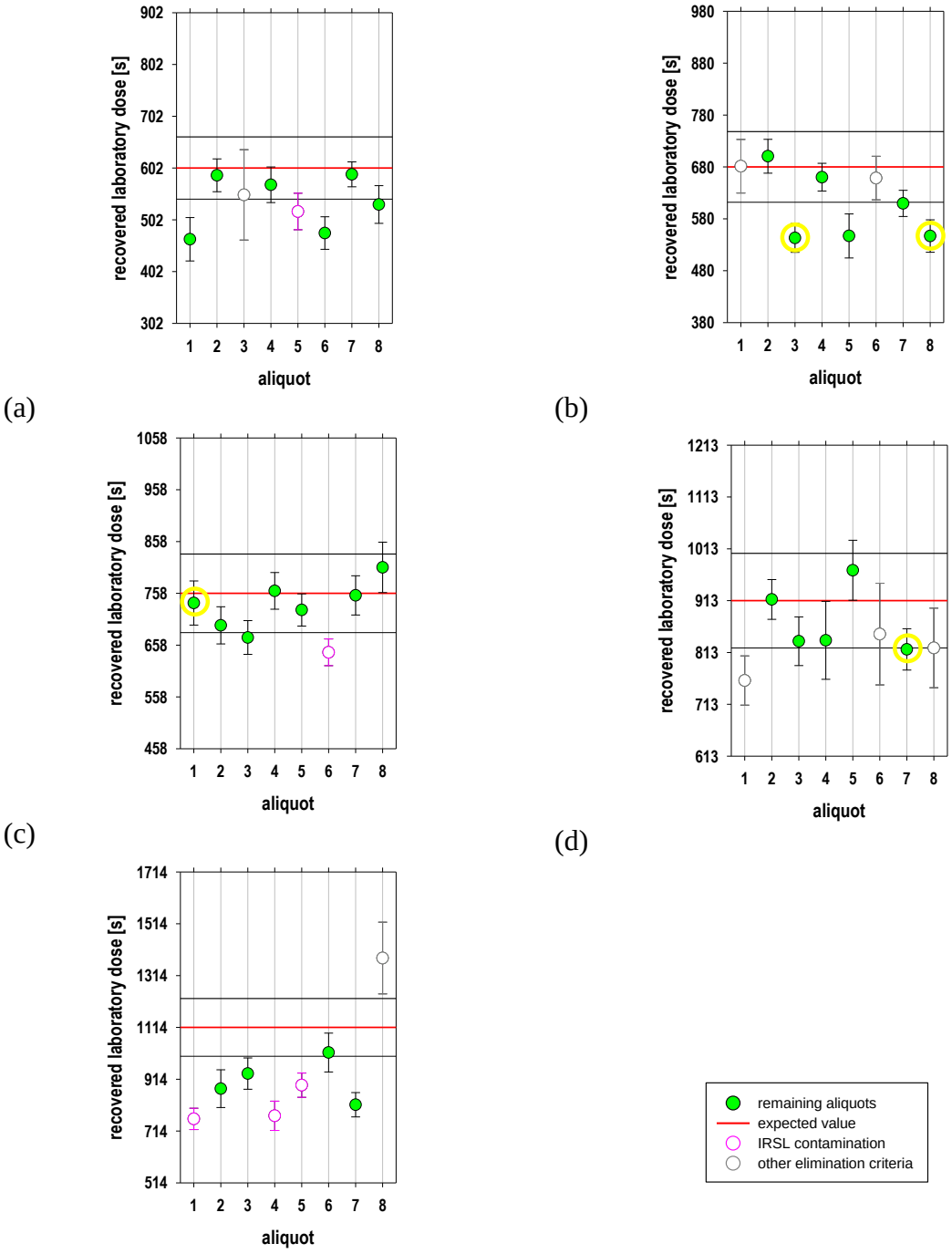


Fig. S-2.6. Dose-recovery test with samples HDS-1660 to HDS-1664, 8 aliquots each. Each sample received a beta-laboratory dose (LAB) the size of which compares to a sample's natural palaeodose. **(a)** HDS-1664 with LAB 602 s (ca. 53.9 Gy), recovered mean with standard deviation and standard error 536.6 ± 55.5 ± 22.7 s. **(b)** HDS-1663 with LAB 680 s (ca. 60.9 Gy), recovered 601.3 ± 67.4 ± 27.5 s. **(c)** HDS-1662 with LAB 758 s (ca. 67.8 Gy), recovered 737.0 ± 44.6 ± 16.9 s. **(d)** HDS-1661 with LAB 913 s (ca. 81.7 Gy), recovered 875.4 ± 65.6 ± 29.3 s. **(e)** HDS-1660 with LAB 1114 s (ca. 99.7 Gy), recovered 911.5 ± 85.8 ± 42.9 s. Although most aliquots meet the expected value (cf. red line) within an error margin of 10 % (cf. black lines), a general trend of LAB being slightly underestimated may be observed. Yellow circles denoting remaining aliquots (green), which do not comply with the pIR-BLSL recycling ratio threshold of 0.9 (alternative / additional test of scrutinizing IRSL contamination; cf. section 2.3.3/supplementary).

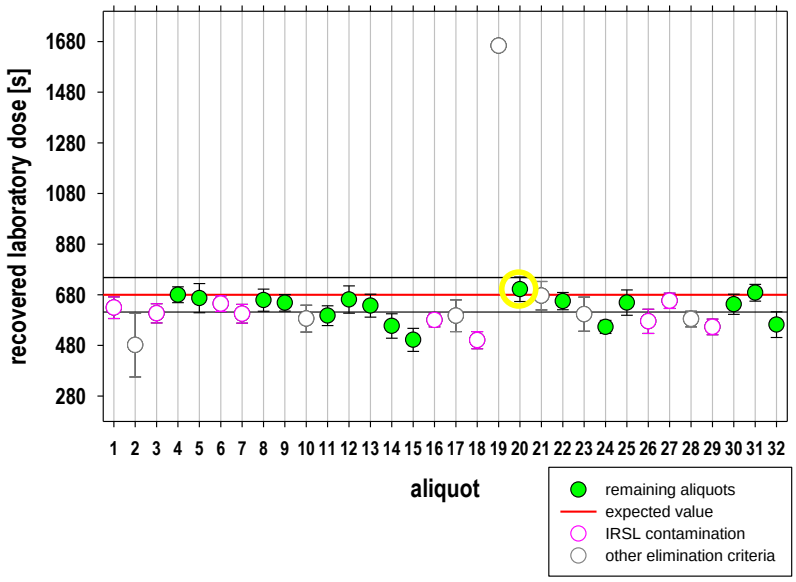


Fig. S-2.7. Dose-recovery test with 32 aliquots of sample HDS-1663 with an administered laboratory dose (LAB) of 680 s (ca. 60.8 Gy). Although most aliquots meet the expected value (cf. red line) within an error margin of 10 % (cf. black lines), LAB is slightly underestimated by ca. 8 % (N = 16; mean with standard deviation and standard error 628.8 ± 57.2 ± 14.3 s; central-age D_e 632.0 ± 13.5 s; overdispersion = 5.5 ± 0.6 %). Yellow circle marking remaining aliquot (green), which does not comply with the pIR-BLSL recycling ratio threshold of 0.9 (alternative / additional test of scrutinizing IRSL contamination; cf. section 2.3.3/supplementary).

2.3.2 Aluminium cups as a possible reason for the slight D_e underestimation

For two other sets of samples, for which we had carried out pretests, datings and additional follow-up dose-recovery tests within the last year, one on reader No. 245 and one on reader No. 240, no such deviations between pretests and follow-up tests had been observed. Therefore, we also examined the different batches of aluminium cups used for the pretests of the Datthausen samples in 2013 and for the follow-up tests in 2016. In autumn 2015 we had observed, with the naked eye, a slight ringlike shade along the outer rim of the bottom side of the aluminium cups, but no actual distortions. While investigating the aluminium cups used for the follow-up tests in 2016 closely under the microscope, we observed a slight (few micron) ringlike bulge at the outer rim of the bottom side of the aluminium cups, indicating that the moulding press is wearing out and needs to be replaced. In contrast to that, previously used and stored aluminium cups (ca. two years old) showed flat bottom sides. It is imaginable that the slight distortion of the recent batch may lead to a slightly less efficient heat transfer to the quartz grains on the aluminium cups. For the heat sensitive (preheat procedure) Datthausen samples this might lead to the observed slight D_e underestimation (cf. Fig. S-2.2 (a) test-1, preaheat-1 of 180 – 200 °C *versus* preheat-1 of 220 – 240 °C). Therefore, we performed a further dose-recovery test on 8 aliquots of sample HDS-1663 with LAB 680 s,

using this time re-used and cleaned aluminium cups of an earlier batch. Yet this test again showed a slight tendency to underestimate the given laboratory dose (Fig. S-2.8).

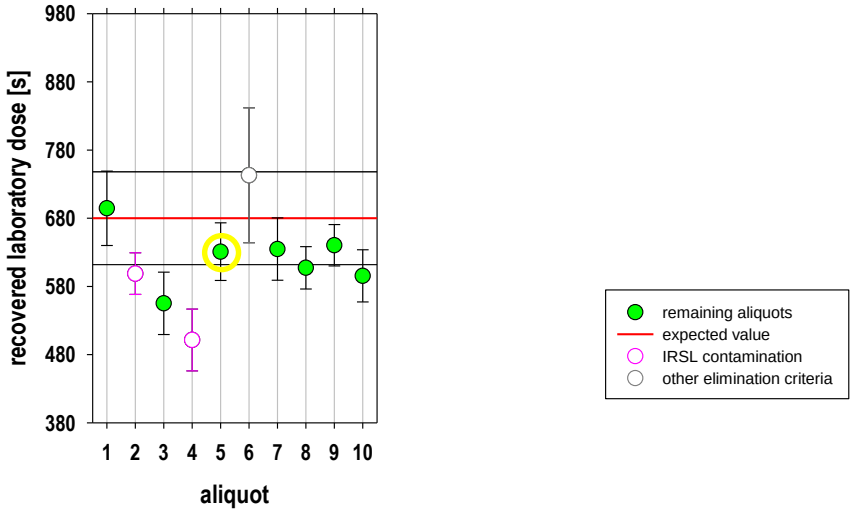


Fig. S-2.8. Dose-recovery test with 8 aliquots of sample HDS-1663 with an administered laboratory dose (LAB) of 680 s (ca. 60.8 Gy). Aliquots prepared on an older batch of aluminium cups. Yellow circle marking remaining aliquot (green), which does not comply with the pIR-BLSL recycling ratio threshold of 0.9 (alternative / additional test of scrutinizing IRSL contamination; cf. section 2.3.3/supplementary). Like for the other dose-recovery tests performed in 2016 the given dose is met within an error margin of 10 %, yet tends to be slightly underestimated (here: ca. 8 %) (N = 6; mean with standard deviation and standard error 622.7 ± 43.2 ± 16.3 s; central-age D_e 622.6 ± 14.7 s; no overdispersion determined by the software Analyst 4.31.7).

2.3.3 Alternative test and elimination of aliquots showing an IRSL signal

Subsequent to the SAR regeneration cycles used to establish the growth curve we had included an extra cycle used for an IRSL test (cf. Table 3/main text, Run 10), which was not considered for D_e determination. However, when re-examining the dose-recovery tests, we also checked the recycling ratio of the BLSL signal of that last regeneration dose point, which had been recorded following an IR-stimulation over 40 s at 125 °C. We observed that some aliquots, which had passed the original IRSL test (< 10 % gross IRSL signal *versus* BLSL signal; cf. threshold criteria in section 3.2.5/main text), showed poor recycling ratios (< 0.9) of the pIR-BLSL signal of the additional regeneration dose point, when compared to the BLSL-signal of an earlier regeneration-dose point which had received the same regeneration dose (70 % of expected D_e, corresponding to Run 4 in Table 3/main text) but was uninfluenced by preceding IRSL readout. Eliminating those aliquots which would deliver poor pIR-BLSL recycling ratios (aliquots marked with a yellow circle in Figs. S-2.5 – S-2.8/supplementary) would for sample HDS-1663 with LAB 1000 s, improve the dose recovery slightly, but not eliminate the overall slight D_e underestimation completely (Fig. S-2.5/supplementary). Nevertheless, we subjected all the dating samples to the same test and eliminated from the data set not only those aliquots with > 10 % gross IRSL/BLSL signal

(conservative IRSL test of main text) but also those which did not comply with a pIR-BLSL recycling ratio of 0.9 – 1.1. This was done to ensure that in our original data sets we had not included possibly fading aliquots which could possibly have led to age underestimation. Afterwards we applied to the remaining data sets the finite mixture model (FMM) and minimum age model (MAM) as detailed in section 3.2.5/main text. The follow-up analyses were performed using the recent version 0.6.0 of the R package 'Luminescence', which due to the further evolution of the software package produces slightly different, but within error margins same results as version 0.3.4 which was originally used for the dating. Due to the lower number of aliquots, age modeling for the follow-up analyses was more restricted than for the original data sets. Thus, e.g., for sample HDS-1660 the two outlier components at the lower end of the D_e distribution could not be detected with a three, but only with a four component model. Yet, apart from these limitations a comparison of the two data sets shows that within error margins they deliver the same results and that the more rigorous elimination of aliquots does not lead to older ages, rendering this procedure as not necessary (Table S-2.2/ supplementary).

Table S-2.2: Comparison of D_es of the data sets remaining after the conservative (main text) and the more rigorous elimination (this section/supplementary) of possibly IRSL-contaminated aliquots. Results are the same within error margins.

	[1] number of aliquots	[1] excluded aliquots	[2] number of aliquots	[2] excluded aliquots	[1] FIN-MIX-MOD sigmab = 0.11						[2] FIN-MIX-MOD sigmab = 0.11							
					Comp-1	Err_Comp-1	Comp-2	Err_Comp-2	Comp-3	Err_Comp-3	Comp-1	Err_Comp-1	Comp-2	Err_Comp-2	Comp-3	Err_Comp-3	3] Comp-4	Err_Comp-4
HDS-1660	27	-	18	-	44.08	5.66	67.70	8.31	**98.94**	**2.76**	44.13	5.73	66.42	13.51	**89.21**	**9.760**	**106.11**	**9.63**
HDS-1661	35	-	28	-	48.72	5.19	**82.31**	**16.98**	82.31	10.13	54.31	11.18	**74.76**	**16.03**	84.00	6.66		
HDS-1662	46	-	40	-	**65.15**	**6.38**	65.15	6.65	89.12	4.26	**66.52**	**3.31**	66.52	2.39	92.36	10.71		
HDS-1663	45	-	36	-	**59.62**	**5.85**	70.81	5.35	113.79	7.79	**57.43**	**4.68**	72.44	3.89	113.99	7.92		
HDS-1664	31	-	23	-	23.67	2.88	**57.10**	**1.87**	85.73	6.55	**56.82**	**2.24**	78.28	7.03	118.49	18.29		

	[1] number of aliquots	[1] excluded aliquots	[2] number of aliquots	[2] excluded aliquots	[1] MAM - without COMP-1 - variant-3 sigmab = 0.11				[2] MAM - without COMP-1 (and COMP-2) - variant-3 sigmab = 0.11				
					minimum dos	Err_neg	Err_pos	Range_min	Range_max	minimum dos	Err_neg	Err_pos	Range_min	Range_max
HDS-1660	27	3	18	2	98.83	7.05	3.08	91.78	101.91	92.12	7.40	7.00	84.72	99.12
HDS-1661	35	2	28	1	82.34	5.09	2.21	77.25	84.55	79.87	4.57	3.46	75.30	83.33
HDS-1662	46	-	40	-	60.78	2.17	6.96	58.61	67.74	66.25	1.95	1.70	64.30	67.95
HDS-1663	45	-	36	-	55.49	2.11	11.35	53.38	66.84	54.94	2.62	8.22	52.32	63.16
HDS-1664	31	1	23	-	49.26	2.52	8.95	46.74	58.21	55.63	8.53	2.99	47.10	58.62

	[1] number of aliquots	[1] excluded aliquots	[2] number of aliquots	[2] excluded aliquots	[1] MAM - without COMP-1 - variant-4 sigmab = 0.20				[2] MAM - without COMP-1 - variant-4 sigmab = 0.20				
					minimum dos	Err_neg	Err_pos	Range_min	Range_max	minimum dos	Err_neg	Err_pos	Range_min	Range_max
HDS-1660	27	3	18	2	99.02	7.72	4.41	91.30	103.43	98.54	5.00	5.26	93.54	103.80
HDS-1661	35	2	28	1	82.42	5.56	3.00	76.86	85.42	81.97	3.25	3.37	78.72	85.34
HDS-1662	46	-	40	-	68.54	4.45	2.18	64.09	70.72	67.85	2.26	2.27	65.59	70.12
HDS-1663	45	-	36	-	60.97	4.17	9.32	56.80	70.29	68.02	3.69	3.08	64.33	71.10
HDS-1664	31	1	23	-	54.44	4.67	9.04	49.77	63.48	60.75	8.78	3.57	51.97	64.32

[1] conservative elimination of IRSL-contaminated aliquots (threshold maximum 10 % gross IRSL of gross BLSL); data processing with R package 'Luminescence' version 0.3.4 (cf. main text)

[2] after eliminating also all aliquots not passing recycling ratio threshold 0.9 - 1.1 of repeated BLSL regeneration dose point subsequent to IRSL-readout (pIR-BLSL); data processing with R package 'Luminescence' version 0.6.0 (cf. this supplementary)

[3] 4 components necessary for HDS-1660 to identify outliers at lower end in data set with only 18 aliquots

2.3.4 Conclusion of the follow-up investigations

The follow-up measurements suggest that sigma b of 0.2 is not a too low value for the Datthausen samples. Applying a more rigorous test of potential IRSL contamination of the quartz separates reduces the set of accepted aliquots but delivers within error margins the same results as the original procedure. Dose recovery tests performed in 2016 tend to slightly underestimate given laboratory doses, which is in contrast to dose recovery tests performed in 2013. A less efficient heat transmission from the heating plate of the luminescence reader to the quartz grains on the aluminium cups during the preheat procedures in 2016 as compared to those in 2013 seems a plausible explanation. The actual reason could yet not be determined in the frame of the here performed follow-up measurements. We would consider it most likely that the dating measurements which occurred in 2013 and 2014 were performed under the same conditions as the dose recovery tests in 2013, which were the base for the dating protocol. Nevertheless, in order to assess the effect of potential D_e underestimations for the dating measurements we added an additional asymmetric (positive) lump error of 10 % on the burial doses (D_bs) and re-calculated the resulting BLSL ages. We indicated this possible additional error in Fig. S-2.9/supplementary (cf. dotted parts of error bars). Yet this measure of precaution would basically not affect the interpretation of the dating results. It would only allow for sample HDS-1660 to also include GIS 8 and GIS 9 as possible periods of soil formation, and for sample HDS-1663 to consider GIS 3 as a further possible period of soil formation.

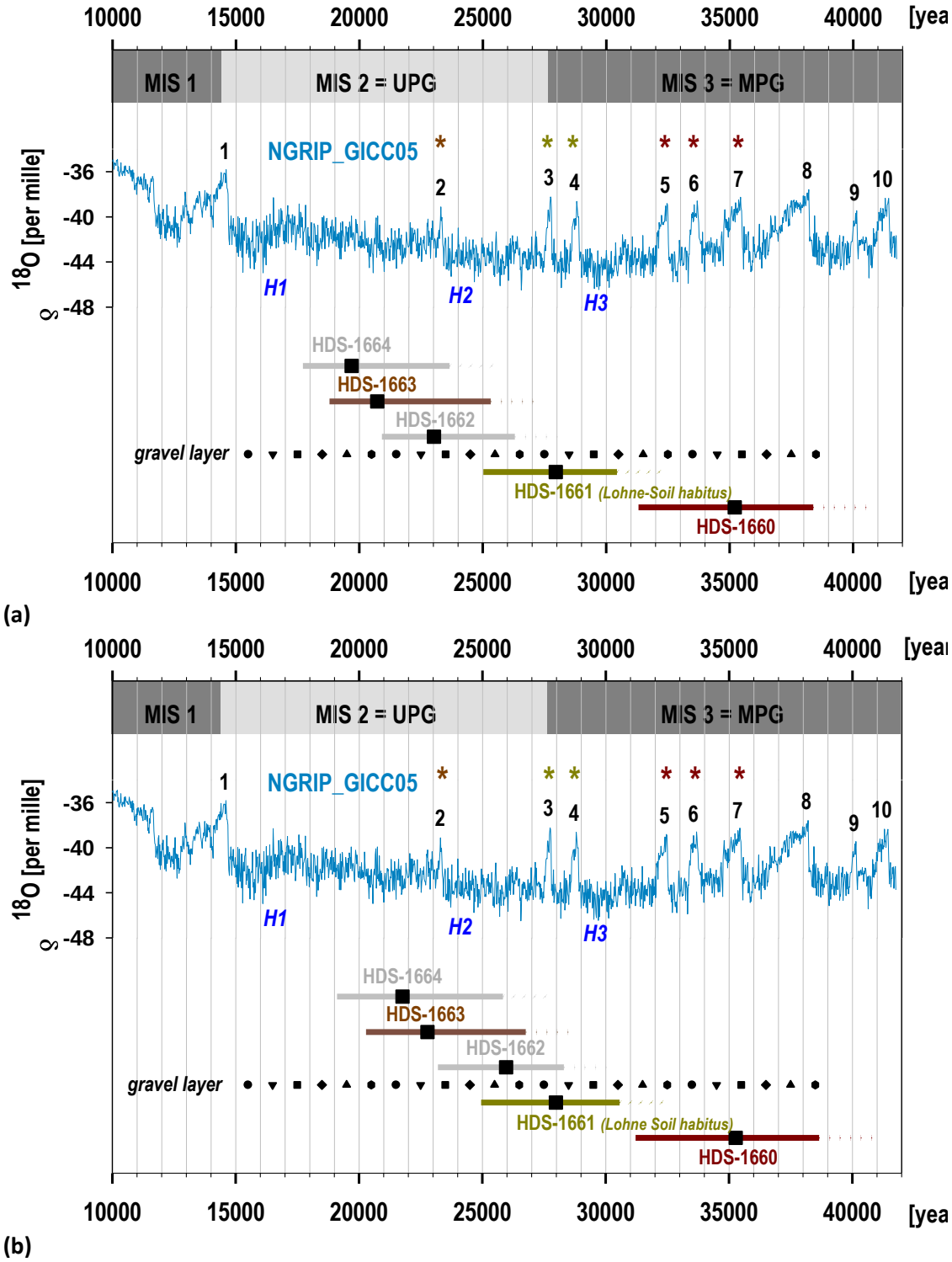


Fig. S-2.9. BLSL-chronology for the loess section at Datthausen reflected in the Greenland Ice Core Chronology 2005 (GICC05). Data from the Centre for Ice and Climate/Niels Bohr Institute/University of Copenhagen: http://www.iceandclimate.nbi.ku.dk/; $\delta^{18}O$-data are based on Dansgaard et al. (1993), GRIP Members (1993) and Johnsen et al. (1997). NGRIP with Greenland Ice Core Chronology 2005 (GICC05) (Andersen et al., 2006; Svensson et al., 2006) released 10 September 2007; file name 'GICC05_NGRIP_20y_10sep2007'; the NGRIP-data for the time-window of 10 to 42 ka are all from core NGRIP2 only. **(a)** MAM approach after elimination of outliers with sigma b = 0.11 (cf. Table 6/ main text, upper part). **(b)** Same as (a), but sigma b = 0.2 (cf. Table 6/main text, lower part). The figure is identical with Fig. 10/main text, but it considers an additional potential asymmetric (positive) lump-error of the burial doses (D_b) of 10 % (dotted parts of error bars).

References

Aitken, M.J., Xie, J., 1992. Optical dating using infrared diodes: young samples. Quaternary Science Reviews, 11, 147–152.

Andersen, K.K., Svensson, A., Johnsen, S.J., Rasmussen, S.O., Biglera, M., Röthlisberger, R., Ruth, U., Siggaard-Andersena, M.-L., Steffensen, J.P., Dahl-Jensen, D., Vinther, B.M., Clausen, H.B., 2006. The Greenland Ice Core Chronology 2005, 15 – 42 ka. Part 1: constructing the time scale. Quaternary Science Reviews, 25, 3246–3257.

Arnold, L. J., Roberts, R. G. 2009. Stochastic modelling of multi-grain equivalent dose (De) distributions: Implications for OSL dating of sediment mixtures. Quaternary Geochronology, 4, 204–230.

Blume, H.-P., Brümmer, G.W., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretschmar, R., Stahr, K., Wilke, B.-M., 2013. Scheffer / Schachtschabel. Lehrbuch der Bodenkunde. 16th edition, 2010, reprint 2013, Spektrum Akademischer Verlag, 569 pp.

Cunningham, A.C., Wallinga, J., 2012. Realizing the potential of fluvial archives using robust OSL chronologies. Quaternary Geochronology, 12, 98–106.

Dansgaard, W., Johnsen, S.J., Clausen, H.B., Dahl-Jensen, D., Gundestrup, N.S., Hammer, C.U., Hvidberg, C.S., Steffensen, J.P., Sveinbjörnsdottir, A.E., Jouzel, J., Bond, G., 1993. Evidence for general instability of past climate from a 250-kyr ice-core record. Nature, 364, 218–220.

Duller, G. (2013): Luminescence Analyst, version 4.10, build 26 March 2013.

Duller, G. (2015): Luminescence Analyst, version 4.31.7, build 04 January 2016.

Greenland Ice-Core Project (GRIP) Members, 1993. Climate instability during the last Glacial period recorded in the GRIP ice core. Nature, 364, 203–208.

Johnsen, S.J., Clausen, H.B., Dansgaard, W., Gundestrup, N., Hammer, C.U., Andersen, U., Andersen, K.K., Hvidberg, C.S., Dahl-Jensen, D., Steffensen, J.P., Shoji, H., Sveinbjornsdottir, A.E., White, J., Jouzel, J., Fisher, D., 1997. The delta O18 record along the Greenland Ice Core Project deep ice core and possible Eemian climate instability. Journal of Geophysical Research, 102 (C12), 26397–26410.

Kadereit, A., Kreutzer, S., 2013. Risø calibration quartz - a challenge for β-source calibration. An applied study with relevance for luminescence dating. Measurement, 46 (7), 2238–2250.

Kreutzer, S., Schmidt, C., Fuchs, M. C., Dietze, M., Fischer, M., Fuchs, M., 2012. Introducing an R package for luminescence dating analysis. Ancient TL, 30 (1), 1–8.

Li, B., Li, S.-H., 2006. Comparison of DeDe estimates using the fast component and the medium component of quartz OSL. Radiation Measurements, 41 (2), 125–136.

Svensson, A., Andersen, K.K., Bigler, M., Clausen, H.B., Dahl-Jensen, D., Davies, S.M., Johnsen, S.J., Muscheler, R., Rasmussen, S.O., Röthlisberger, R., Steffensen, J.P., Vinther, B.M., 2006. The Greenland Ice Core Chronology 2005, 15–42 ka. Part 2: Comparison to other records. Quaternary Science Reviews 25 (24), 3258–3267 (Shackleton special issue).

Timar, A., Vandenberghe, D., Panaiotu, C. E., Panaiotu, C. G., Necula, C., Cosma, C., van den Haute, P., 2010. Optical dating of Romanian loess using fine-grained quartz. Quaternary Geochronology, 5, 143–148.

Wallinga, J., Murray, A., Duller, G., 2000. Underestimation of equivalent dose in single-aliquot optical dating of feldspars caused by preheating. Radiation Measurements, 32, 691–695.

Wintle, A.G., 1973. Anomalous fading of thermoluminescence in mineral samples. Nature, 245, 143–144.