This manuscript presents the results of an LSM application using two different approaches according to BC and VG to reproduce soil water mass, volumetric water content, and drainage water flux volume observed in seven lysimeters over a period of more than five years. Furthermore, approaches by Braud et al. (1995) and Valiantzas (2011) are tested to simulate the soil hydrology of these lysimeters. They derived hydrodynamic parameters directly from the observation and compare them with several pedotransfer functions commonly used by LSMs. The LSM used has a multilayer diffusion approach of the Interaction-Soil-Biosphere-Atmosphere (ISBA) model, which solves a variant of the Richards equation.

- The term "drainage" in this context means the transit of a liquid through a porous medium. In the present case, it is water through the upper soil layers. It is neither a quantity nor a volume. Therefore, if the amount of water is to be addressed this must be explicitly stated as drainage water.

Thank you for this comment. We agree with you and have agreed to change the term “drainage” by “drainage water”.

- For the lysimeters, the experimental setup is sufficiently described, but the lower boundary condition is not mentioned in detail as a special feature of the lysimeter. Since drainage in particular is considered as a special aspect, this has to be described in detail for the lysimeters. Especially the consequences/impacts of the chosen design on the drainage amount of water must be discussed. Otherwise, it is assumed here that the lower boundary layer of the lysimeter corresponds to a naturally layered soil, and this is de facto not the case.

Yes, some information on the setting of the simulation was missing.

In this study, the soil discretization was adapted to the lysimeter depth and to the measurement depths: 13 soil layers are used, with nodes at 0.01, 0.04, 0.1, 0.2, 0.4, 0.6, 1.0, 1.2, 1.4, 1.6,1.8 and 2 m. A free drainage condition is used at the bottom of the soil column.

Such conditions imply that the bottom of the soil should be saturated to generate. You’ll see below that this can be observed, at the answer to your question on Line 181.

Additionally, a one-year simulation is performed and the output of the simulation are used

- The methods section on the comparison between the model predictions and the lysimeter observations is very unclearly written and needs a more comprehensible description.

Thank you for these useful comments. Indeed, we did not provide details on the soil parameters calibration procedure, as we mainly focused on the difference between the soil parameters derived from in-situ data to those derived from pedotransfer functions. We will add the details in the revised version of the article.

The calibration of the soil parameters was performed using two methods: i) an objective least squares function which minimizes the sum of the squares of the deviations and corresponds to maximizing the likelihood with a normal distribution (function nls of rstudio). ii) the package SoilHyP (Dettmann et al., 2022) which uses the Shuffled Complex Evolution (SCE) optimization.

The two methods converge on very closed estimated parameters ($\delta b: 1.6;\ \delta n: 0.01; \delta wsat: 0.0085\ (m^3 m^{-3}); \delta\alpha: 2\ (m^{-1});\ \delta\Psi sat: 0.019\ (m)$).

Statistical scores and errors metrics on the estimation of the soil water retention curves (WRC) are presented in Table R2.1 for the BC66 and VG80 relationships at each depth. On the entire fit and close to saturation ($> \omega_{sat} * 0.9$), values median, minimal and maximal are presented. The fitted are generally better in depth, and a better r² for VG80 than for BC66 near to saturation. Near saturation, the NRMSE is also better for VG80 than BC66.

Table R2.1 : Statistical scores (regression, r²) and errors metrics (Normalized Root Mean Square Error, NRMSE, and Akaike information criterion, AIC) for the BC66 and VG80 relationships at each depth (20-50-100 and 150 cm) on the total entire fit (Total) and close to saturation ($> \omega_{sat} * 0.9$) for all lysimeters for soil water retention curves (WRC).

	Total					
	r²		NRMSE		AIC	
	BC66	VG80	BC66	VG80	BC66	VG80
	20cm					
median	0,891	0,886	0,337	0,34	1755,675	1777,015
min	0,846	0,84	0,229	0,189	23,154	34,243
max	0,958	0,966	0,393	0,401	5124,888	4938,271
	50cm					
median	0,69	0,84	0,67	0,4295	931,76	550,886
min	0,024	0,666	0,421	0,156	-1045,18	-1020,23
max	0,897	0,976	1,761	0,693	5828,86	5773,81
	100cm					
median	0,968	0,967	0,2285	0,1975	-2003,996	-2061,039
min	0,691	0,684	0,111	0,095	-3454,685	-3501,568
max	0,989	0,991	0,624	0,563	2368,327	2270,159
	150cm					
median	0,966	0,97	0,244	0,2435	-1842,222	-1546,381
min	0,711	0,66	0,122	0,159	-5659,271	-5651,963
max	0,991	0,987	0,608	0,7	6884,241	6852,803

	Close to Saturation					
	r²		NRMSE		AIC	
	BC66	VG80	BC66	VG80	BC66	VG80
	20cm					
median	0,98	0,99	1,136	0,757	-129,256	-216,541
min	0,967	0,978	1,11	0,665	-249,818	-419,598
max	0,983	0,995	1,212	1,7	26,347	4,815
	50cm					
median	0,938	0,926	0,629	0,483	220,637	-150,892
min	0,739	0,75	0,383	0,369	-105,151	-868,689
max	0,99	0,999	0,775	1,041	2154,151	1429,798
	100cm					
median	0,932	0,953	0,493	0,568	111,815	-84,368
min	0,739	0,773	0,219	0,161	-118,14	-226,769
max	0,979	0,987	1,12	1,412	770,041	642,032
	150cm					
median	0,849	0,931	0,403	0,313	247,332	198,666
min	0,543	0,67	0,187	0,17	-209,217	-2331,309
max	0,964	0,996	0,598	1,02	720	696,064

The calibration of the hydraulic conductivity curves (HCC) was also realized. For each depth, we estimated the hydraulic conductivity at saturation with the water volumetric content and the drainage at 2m for at least 3 years by averaging the drainage values for each water volumetric content value. Statistical scores are shown in Table R2.2 for the four relations. RMSE are better at depth with closer differences than at surface, and always better for VGBC (<0.5 mm/h). AIC are relatively closed between the four relations.

Table R2.2 : Statistical scores (regression, r²) and errors metrics (Root Mean Square Error, RMSE, and Akaike information criterion,AIC) for the BC66 and VG80 relationships at each depth (20-50-100 and 150 cm) for calibration for hydraulic conductivity curve (HCC).

	r^2				RMSE (mm/h)				AIC			
	BC66	VG80	VGc	VGBC	BC66	VG80	VGc	VGBC	BC66	VG80	VGc	VGBC
	20cm											
median	**0,734**	0,252	0,362	0,475	0,545	0,626	1,201	**0,467**	-20,406	**-230,292**	92,441	-35,782
min	0,318	0,214	0,153	0,242	0,476	0,546	0,604	0,427	-34,891	-272,773	-28,618	-67,644
max	0,766	0,361	0,415	0,482	0,928	1,008	5,049	0,767	39,966	-24,493	178,235	-32,912
	50cm											
median	**0,601**	0,295	0,415	0,332	0,669	0,504	0,758	**0,426**	**-27,32**	-27,592	-26,061	-25,694
min	0,346	0,106	0,181	0,185	0,199	0,223	0,345	0,132	-37,285	-62,605	-36,637	-46,585
max	0,846	0,524	0,712	0,653	0,82	0,788	9,66	0,708	65,976	-12,016	226,709	-11,994
	100cm											
median	**0,688**	0,417	0,578	0,36	0,352	0,425	0,407	**0,3**	-23,813	-25,735	-23,651	**-30,998**
min	0,398	0,125	0,208	0,19	0,283	0,366	0,352	0,187	-63,182	-255,385	-152,687	-65,784
max	0,808	0,876	0,876	0,888	0,649	0,659	0,809	2,034	51,406	-16,898	65,95	142,02
	150cm											
median	**0,599**	0,248	0,472	0,307	0,548	0,659	0,612	**0,485**	**-34,007**	-27,345	-30,193	-32,703
min	0,194	0,132	0,155	0,184	0,298	0,366	0,357	0,175	-59,104	-468,014	-32,471	-88,413
max	0,96	0,584	0,801	0,667	0,747	0,742	0,946	0,675	59,053	-17,797	81,5	-21,512

- Why these applied models were selected is not convincingly presented, especially since there are more current modeling approaches that promise better simulation of processes and results.

The model ISBA-SURFEX solves both the water and energy budgets at a fine temporal scale so as to provide atmospheric models with surface flux (latent and heat fluxes) and boundary conditions (surface temperature and albedo for instance). It is used in regional hydrological models, as well as in weather forecast and climate models. Therefore, it is important to assess its simulations of the soil water flux.

Like many soil–vegetation– atmosphere–transfer schemes (SVAT), it uses the relation of BC66. Compared to data from lysimeters, it is obvious that this relationship is not the most correct to reproduce WRCs close to saturation. This is why we decided to use the relations of VG80, but the hydraulic conductivity curves of VG80 requires new parameters, and VG80 is unstable close to small values of n.

The BCVG relation is one solution to solve this numerical problem; an approach has been used in other studies (Braud et al., 1995, Valiantzas, 2011), which consists of combining VG80 water retention curves and BC66 hydraulic conductivity curves. We are well aware that BCVG seems not very physical from a theoretical point of view, but from a pragmatic and statistical point of view, water retention curves and HCC can reproduce the observations on lysimeters very well. The advantage of this approach is also the limited number of parameters and can be spatialized easily at regional or global scale in the ISBA surface model. Following the comments of the 3st reviewer, Wolfgand Durner, we also tested an intermediate approach by truncating the pore-size distribution (Iden et al., 2015), which provide promising results (see answer to reviewer 3)

- Lysimeters provide "point" information compared to LSM. Here, indications are missing how this discrepancy is addressed or how lysimeter results could be scaled.

The ISBA LSM model is applied at global, regional and local scales. Several studies showed the good performance of ISBA at local scale : Boone et al., 2000, Calvet et al., 1999, Decharme et al., 2011. Moreover, studying lysimeters with LSM allows us to verify and to propose improvement directions for simulations at larger scale, such as the integration of a heterogeneous profile with depth.

Line 107: A specification of the measurement resolution is missing here.

On the GISFI site, TDR probes are RIME-PICO32 sensors with internal TDR-electronics. They are set horizontally and record the water content in cm3 cm− 3 (± 0.01) on an hourly basis. The calibration was performed on two measurements, one in dry and one in water-saturated condition. In the OPE site, soil moisture sensors used (UMP-1Umwelt Geräte Technik GmbH) are based on frequency domain reflectometry (FDR) method and measure local change in dielectric permittivity.

Tipping bucket resolution is 0.1mm.h-1 on the two sites.

Line 181 ff: Is this the case? It is often stated that in zero-tension lysimeters, the seepage water formation takes place under water-saturated conditions. I am not aware of any study that has decisively investigated this. Of course, small-scale saturated structures are also conceivable with corresponding fingering. Is this the case? Especially before the background of a very heterogeneous material of a former industrial site, which was filled manually into lysimeters. Hydrophobic structures are also conceivable.

Yes it is the case. In the Figure R2.1, variation of the total hydraulic pressure head between 50 and 150 cm ($\frac{dH}{dz}$) of depth is represented for each lysimeter. Lines blue and red represent a gradient equal to 1 and 0 respectively. It can be observed that the gradient is in most cases equal or very close to 1. Some variations exist, in particular for lysimeter G3, because of strong and deep roots.

Nevertheless, when the saturation of volumetric content in the soil is reached, the gradient is equal to 1, and we can assume at this moment, ksat is equal to observed drainage. The lower boundary condition is a free drainage.

The process of hydrophobicity is negligible in these lysimeters.

.

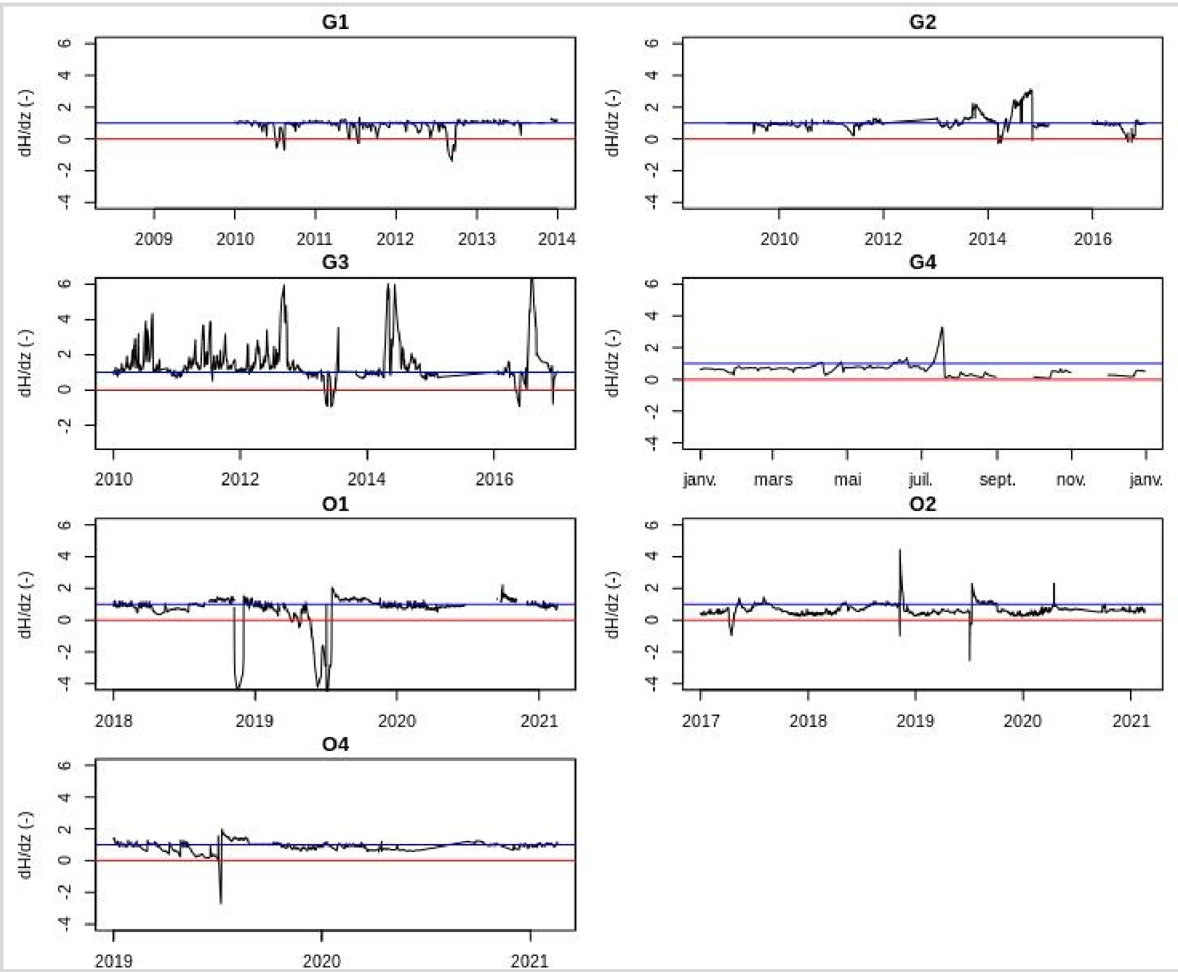

Figure R2.1 : Variation of the total hydraulic pressure head between 50 and 150 cm ($\frac{dH}{dz}$) of depth is represented for each lysimeter. Lines blue and red represent a gradient equal to 1 and 0 respectively

Line 200 ff: Should be discussed later, because it is manual filling with disturbed profiles. This has an impact on the parameter estimation.

We propose to add in the discussion this point :

Our observations show that the soil hydrodynamic parameters in each lysimeter are strongly heterogeneous with depth, while LSMs generally use homogeneous profiles. Lysimeters were filled with preserved soil columns at the OPE, and manually at the GISFI. The fact that the lysimeters are filled in a manual way can explain this evolution, although they have been filled to preserve the bulk density of the soil. Factors like compaction and structuration can be present on lysimeters (Durner, 1994, Séré et al., 2012).

Using additional experiments with such homogeneous profiles, we found that even if the simulated soil water and drainage dynamics remain acceptable compared to the observations, all the skill scores are worsen compared to the experiments with a heterogeneous profile.

Line 233 ff: “At the bottom of the soil“ What do you mean by this? This wording is very unclear or does not make sense.

We meant that the drainage is measured at 2 m of depth, the depth of the lysimeters. We’ll make it clearer in the revised version.

Line 242: Masses of what? Water?

Yes, it is well the total water mass on lysimeter.

We proposed to modify the article like this : 4.1.1 “Total Water Mass”.

Line 266f: I do not understand this argumentation. The temporal resolution is criticized as limiting and therefore I reduce the temporal resolution even more or aggregate the data?

We choose to aggregate the data on daily drainage to not be limited by the threshold resolution of the measurements, since at the hourly time steps, the measurements are most often 0 or 0.1 mm…. Aggregating at a daily time step provides observations that are easier to interpret .

Line 265ff: In order to be able to classify the different seepage water quantities, a distinction must be made between vegetated and unvegetated lysimeters. This has been done. But to be able to investigate or classify the differences between the vegetated lysimeters, measurements of the crop development (LAI) or the crop yield (harvest amount), etc. are absolutely needed. Only with this information different ETp results can be classified.

We have no information of the crop yield for these lysimeters, notably because vegetation were not harvested (except for G4). No LAI measurements were available on these lysimeters, but LAI is simulated (line 220-225). For more precision, we propose to add a figure with a mean annual cycle of the LAI for lysimeters with vegetation and the daily simulated LAI in the supplementary data (Figures R2.2 and R2.3.

Moreover, the water budget of each lysimeter is also estimated (Figure 8 in the paper) and can give an information of the ETp.

.

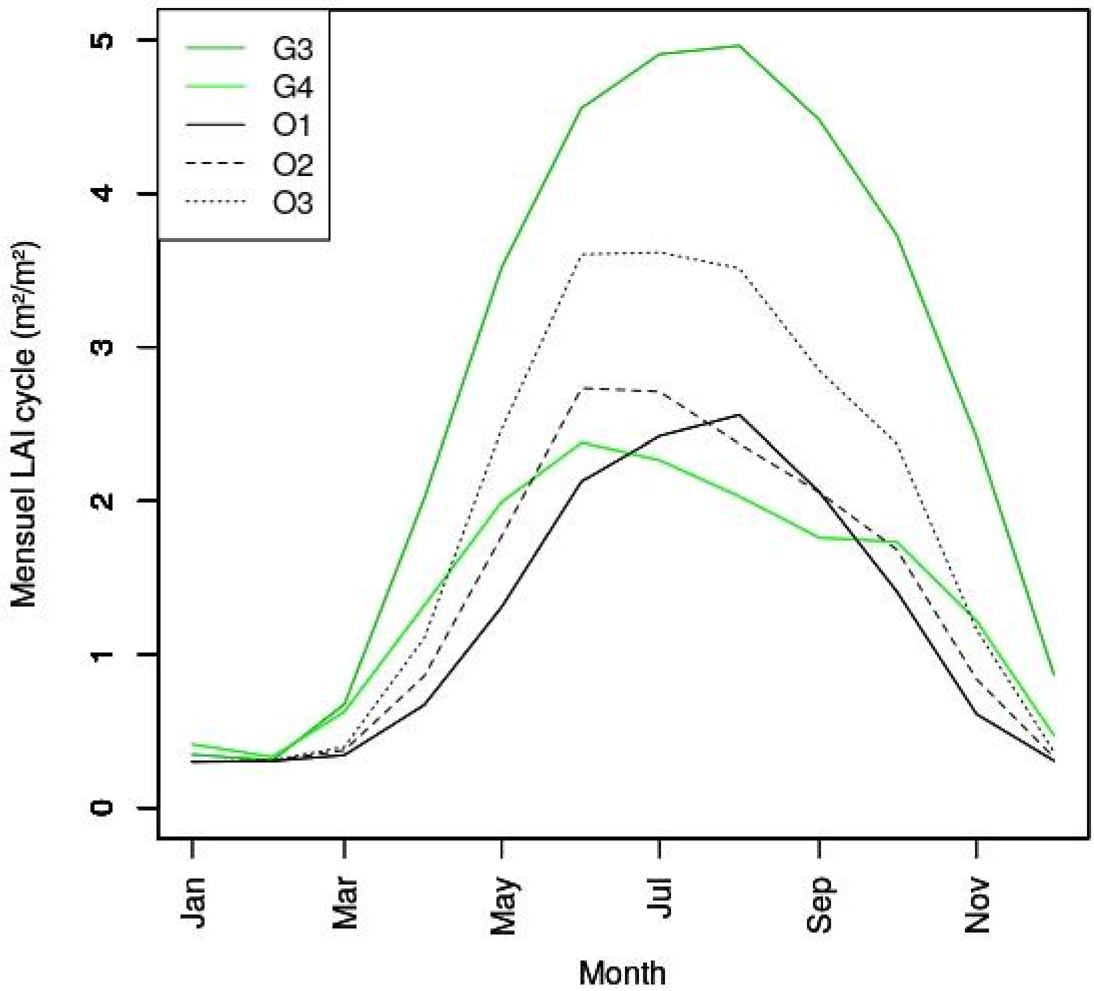

Figure R2.2 : Mean LAI cycle simulated for lysimeters for each lysimeter

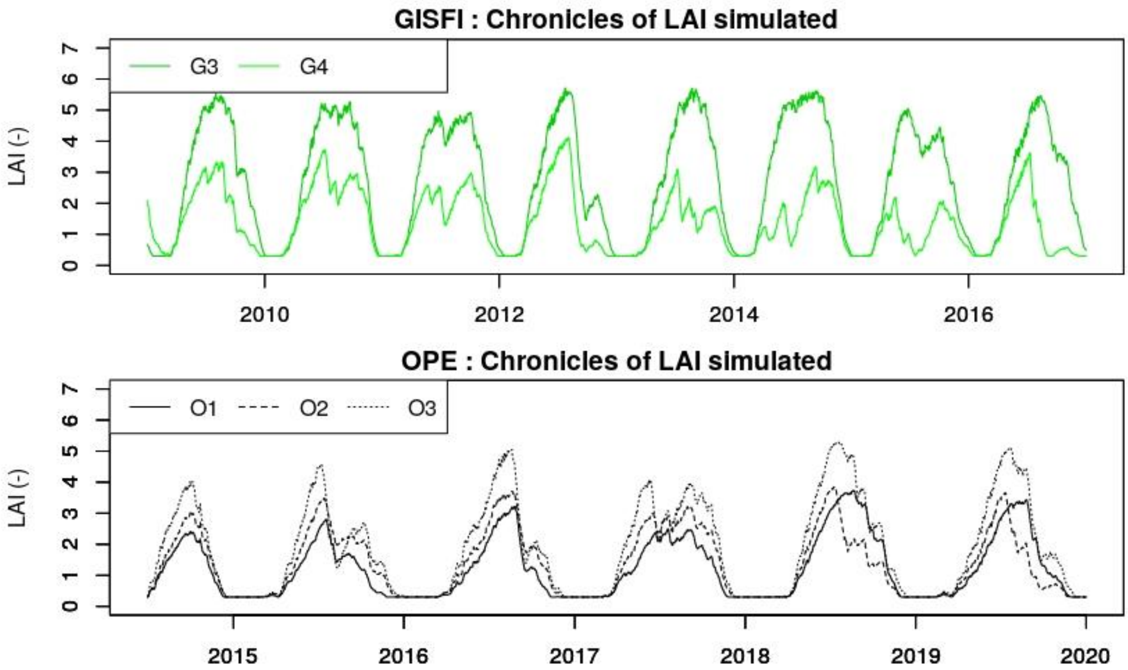

Figure R2.3 : Daily evolution of the LAI simulated for each vegetated lysimeter

Line 289 ff: Also, non rainfall water like dew, hoar frost, etc.

Thank you for this comment. We specified this water budget is conducted on an annual scale and dew/rime is negligible at this scale, although dew deposition is simulated by the model, and is considered in the balance as negative evaporation. Also, there was no freezing/irrigation on these sites during the measurements.

Table 1: Regarding the contents of the table: am I correct in assuming that the remainder is 100% silt? If not, what then? But should still be presented in more detail for clarity. To standardize the presentation, the number of decimal places should be the same for all data.

Yes, you are right. Table 1 indicates the proportions of clay and sand. For example, G4 lysimeter at 20 cm, is composed of 32% clay and 25% of sand, the remainder is the proportion of silt (43% here). We will modify this table by adding the proportion in silt as well as putting the same number of decimal for more comprehension.

Table R2.3

Table 1. Description of lysimeters: filling method, soil type, vegetation cover, number of texture observations, and textures (in % of clay, sand ans silt) at different depths. HyCa stands for Hypereutric Cambisol

Site	GISFI experimental station				OPE experimental station		
Lysimeters	G1	G2	G3	G4	O1	O2	O3
	Fill	Fill	Fill	Monolith	Monolith	Monolith	Monolith
Soil	Technosol	Technosol	Technosol	Cambisol	HyCa	Cambisol	HyCa
Soil cover	bare soil	bare soil	Alfalfa	Grass	Grass	Grass	Grass
Layers	1	1	1	4	6	1	6
Bulk density (kg.m^{-3})	1300	1300	1300	1300	1700	1700	1700
USDA soil type	Sandy Clay Loam	Sandy Clay Loam	Sandy Clay Loam	Clay Loam	Clay	Clay Loam	Silty Clay Loam
(Sand,Clay,Silt) (%,%,%)							
Homogeneous	(61.6, 14.3, 24.1)	(61.6, 14.3, 24.1)	(62.4, 15.2, 22.4)	(32.0, 25.0, 43.0)	(3.0, 36.0, 61.0)	(31.0, 41.0, 28.0)	(18.0, 47.0, 36.0)
0.2m	""	""	""	(20.0, 15.0, 75.0)	(11.0, 4.0, 85.0)	(50.4, 18.0, 31.6)	(24.0, 28.0, 48.0)
0.5m	""	""	""	(17.0, 26.0, 57.0)	(0.0, 67.0, 37.0)	(42.0, 27.0, 31.0)	(16.0, 53.0, 31.0)
1m	""	""	""	(34.0, 33.0, 33.0)	(0.0, 19.0, 81.0)	(22.0, 6.0, 72.0)	(16.0, 53.0, 31.0)
1.5m	""	""	""	(56.0, 24.0, 20.0)	(0.0, 19.0, 81.0)	(22.0, 6.0, 72.0)	(16.0, 53.0, 31.0)

Figure 3, 5-7: Measurement units of the Y-axes are missing

The units were provided in the caption but not in the figures. We will add the units in the plots for a better understanding.

Furthermore, I share the remarks of the reviewer 1.

Durner, W.: Hydraulic conductivity estimation for soils with heterogeneous pore structure, Water resources research, 30, 211–223, 1994.

Séré, G., Ouvrard, S., Schwartz, C., Pey, B., and Morel, J.-L.: Identification of hydric functioning patterns during the early pedogenesis of a Constructed Technosol, in: 1. International Conference and Exploratory Workshop on Soil Architecture and Physico-chemical Functions "CESAR", pp. 426–p, Aalborg University; Faculty of Agricultural Sciences, Aarhus University, 2010.