I sincerely thank Gerrit de Rooij for his thoughtful and detailed assessment of our work. I am sure this consumed a considerable amount of time and this investment is highly appreciated. Gerrit de Rooij's critique is mainly targeted towards sections 1 and 2 (and a little about section 3). Before I address those points in detail I would like to clarify a few issues dealing with the usefulness of the proposed approach, its predictive potential, present modelling evidence (including spatial averaging).

The added value of our approach

Figure 1 shows the time series of soil moisture observations observed in the Colpach (blue lines) and the Wollefsbach (red lines). It is interesting to see that the ranks of quite a few sensors are stable in time. Why? Because their spatial variability is not purely random! Or course the soils in both areas are not homogeneous, hence these differences cannot be attributed to HAND differences alone (i.e. perturbations around different equilibria). I guess we see a mixture of HAND and variability in texture.

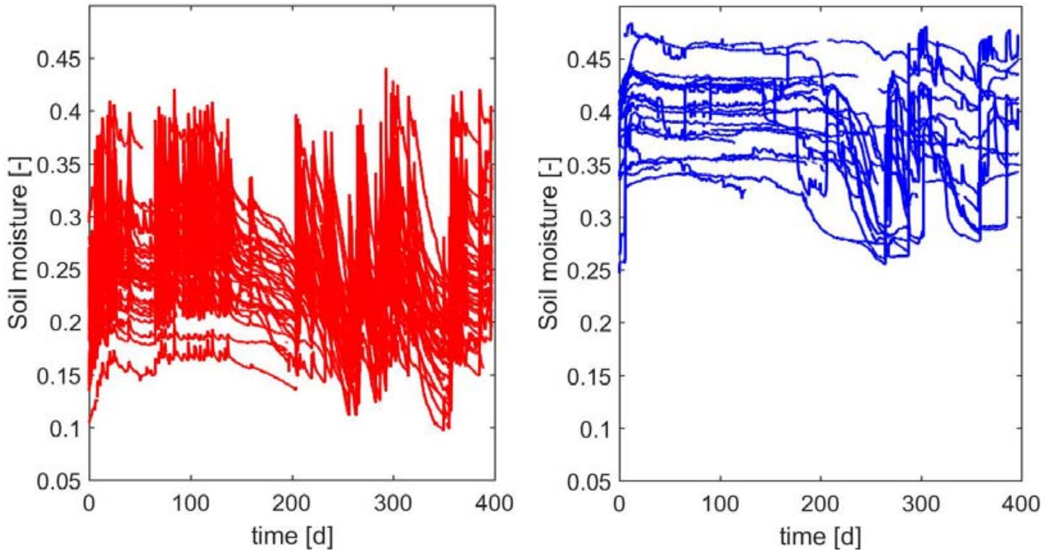


Figure 1: Soil moisture time series in the Colpach (left panel) and the Wollefsbach (right panel).

Despite of this interesting finding a comparison between both samples appears a little messy, and one would again conclude that the Wollefsbach is wetter than the Colpach. The comparison we presented Figure 8 of our manuscript, corroborates that quite a bit of this variability among the different sites is indeed systematic and explainable from the energetic point of few. And again we see that the Colpach operates in the P-Regime and while the Wollefsbach drops into the C-Regime.

Does it help to explain runoff generation in the catchment?

This was asked by Thilo Streck when I discussed the approach recently on a workshop of the German Soil Science Society. While Thilo found the approach "interesting" he pointed out that runoff response to rainfall is not generated everywhere the catchment, which means that the energy state at remote locations might be pretty unimportant for this. This is of course very true and today we know that during most of the rainfall events, runoff is mainly produced in the riparian zone (unless it becomes extreme and the hillslope switch on). Figure 2 shows the free energy state of an observation site which is located in the riparian zone of the Colpach (there are unfortunately only 2 sites).

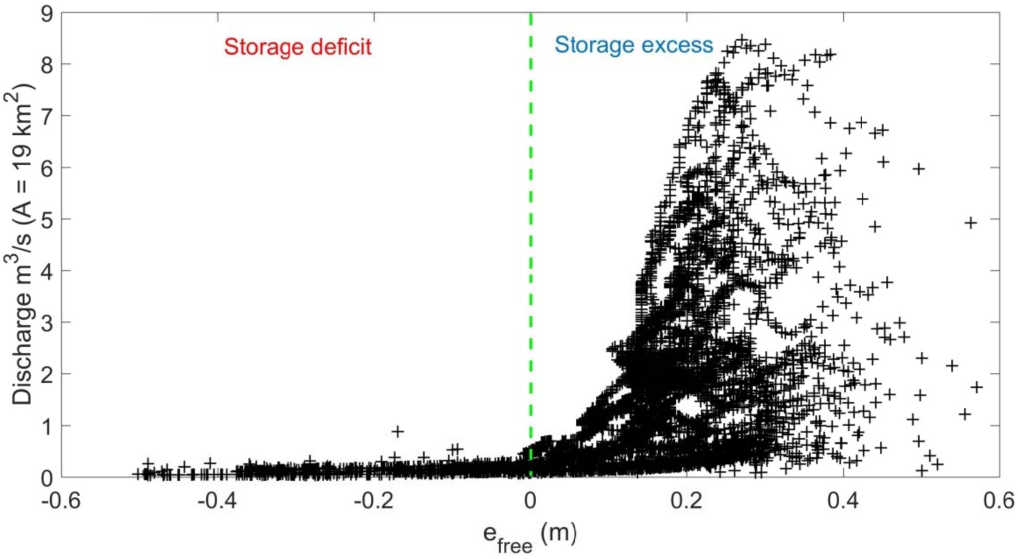


Figure 2: Discharge at gauge Colpach (19.km^2) plotted against the energy state of soil moisture in the riparian zone

Note the threshold character: discharge for efree < 0 (storage deficit) is small while it shows a enormous spread when efree > 0. The spreading reflects the differences in rainfall forcing. I think this plot corroborates that the energy state of the soil water content in the riperian zone is a nice predictor to characterise the onset of rainfall runoff production in the riperian zone and it corroborates that the distance to the equilibrium provides valuable information, and provides a theoretical background to predict the onset of enhanced runoff production.

Modelling evidence

Figure 3 compares the energy states of soil moisture observed that the two selected sites discussed in the manuscript with those from simulations with a 2 dimensional physically based model (based on the Richards equation and more).

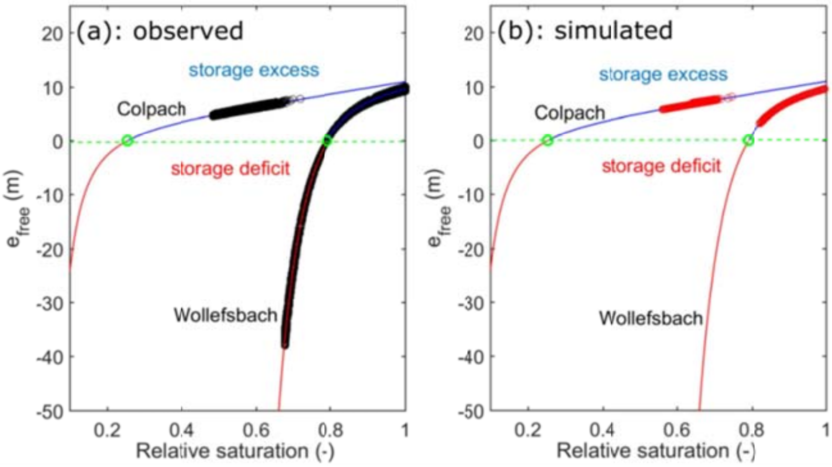


Figure 3: Free energy of observed (Panel a) and simulated (Panel b) soil moisture time series at the selected sites in the Colpach and the Wollefsbach catchment.

The model has been parameterized for both catchments using extensive data and it was shown to predict the water balance and runoff dynamics of at least the Colpach very well as explained in Loritz et al. (2017). Please note that for the Colpach simulations and observations are in a good accordance (the mismatch is of gradual nature), while simulations cannot reproduce the drop into the C-Regime during the summer period in the Wollefsbach. This is because the model systematically overestimates soil moisture in these soils during dry spells. I think this is a nice example that the approach is helpful to discriminate substantial from gradual model errors.

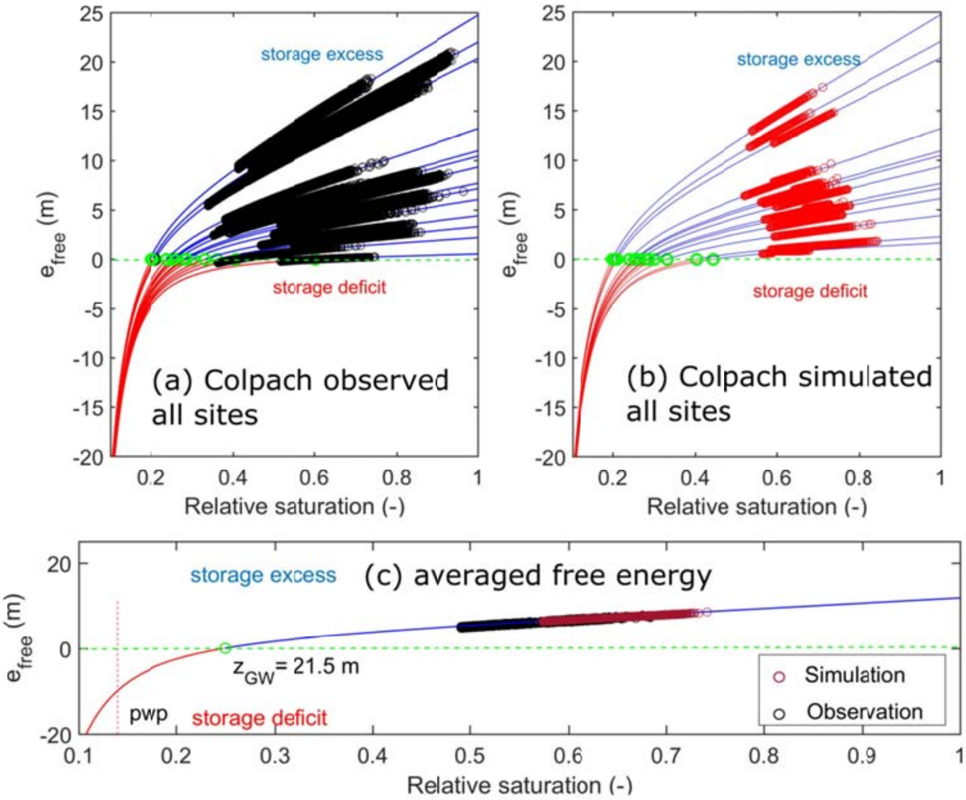


Figure 4: Observed and simulated free energy of the soil water stock in their corresponding energy level function (Panel a and b). Panel c shows the corresponding dynamics of the averaged simulated and observed free energies as function of saturation.

Figure 4 panel a) and b) compare observed against the simulated free energy states of the soil water stocks in energy level functions of the Colpach. Observations spread across a much wider range than the simulations. This mismatch is easily explained by the facts that the model setup of Loritz et al. (2017) does neither account for small scale soil heterogeneity, nor for larger scale variability of rainfall or of forest vegetation. We tested whether the model setup is at least capable to predict the average free energy state of the observed soil water stock. To this end we averaged the free energy states of the observed soil moisture time series and plotted those as a function of average soil saturation into the corresponding average energy state curve (Figure 4c, black circles) and compared those deviations to the average simulated free energy states (Figure 4c brown circles). Though the matching is much better for the averaged dynamics, the simulation shows a slight positive bias of 0.6 m ($5.9*10^3$ Jm^{-3}).

The presented model evidence underpins a) the general idea discussed in the manuscript well and b) that, due to the additive nature of energy, an inter-comparison of averaged energy states in the averaged energy state function is straightforward and helpful.

In summary think that the above presented figures and arguments are helpful to underpin the value of proposed perspective.

In the following I'll address the points brought up by Gerrit de Rooij. Before I go into the details I'd like to stress that we indeed missed to refer to relevant literature, including Gerrit's work. This was not done on purpose and we will happily correct for this in the revised manuscript. I will also not repeat our reply to reviewer 1 in respect to the quote of Aristotle and the language issue here. I guess this goes pretty much into the direction of what Gerrit de Rooij would expect. Last not least I' d like to share that I learnt my thermodynamics from a different sample of textbooks, which are usually used in theoretical physics (Honerkamp and Roemer) – I am usually a little hesitant to refer to those in hydrological papers, but this explains maybe our different background with respect to thermodynamic textbooks. Now let me come to the details.

Why Helmholtz instead of Gibbs free energy (and our understanding of potential energy).

This is because I thought that it doesn't make too much of a difference. Thermodynamics provides a large set of thermodynamic potentials: inner energy, free energy, enthalpy and free enthalpy (the terms I learned in German). At the end of the day they all have the same current (Joule) and their respective use is strongly related to thermodynamic processes in connection with "reservoirs" and to the set of state variable which are controlled during these processes.

U is inner energy and dU =TdS (heat) –PdV (mechanical work) + μdN (chemical energy, where N is number of molecules in the gas and μ chemical potential, alternatively we can use the mass M, express μ not as energy per molecule but per mass. Note this is not a total differential, as changes in T, p and μ are not considered). Gerrit is correct that the way from U to one of the other potentials is by the Legendre transformations he showed. This relates to the degrees of freedom and what is controlled in the system. Helmholtz free energy is E=U-TS, and assumes that the system is not in contact with a heat reservoir, hence temperature is not controlled.

dE=-SdT (also heat)-PdV (mechanical work as defined by Boltzmann) + μdN (chemical energy, or molar free energy after Bolt and Frissel). We preferred this because we treat the soil as rigid. Hence dVsoil = 0 which means we can neglect mechanical work. We agree that if we select G we end up with dG=-SdT (heat)-VdP (I learned this is also work) + μdN, the latter might be interesting in the aquifer. So maybe this is the next step when looking at both the aquifer and the unsaturated zone. I admit that Bolt and Frissel account for an additional work term (which relates to the scalar produce of forces and displacements).

So if we focus on the change of free energy associated with changes in soil water content (and neglect their effect on soil temperature and use the mass related formulation we end with dE = μdM. One way to relate this term to the matric potential, is to relate it to chemical energy (as proposed by Kleidon and Schimansky (2008) or Hildebrandt et al. (2017)). I come back to this below.

Potential energy does not belong to inner energy and I also think that it is misleading to relate the matric potential (which I like very much) to potential energy. To my understanding potential energy relates to the position of an object/body in the gravity field (which does not involve any contact of bodies or fluids) or alternatively an electrical charge in an electrostatic field, or a Hadron in the potential well of the nucleus. I think relation to an elementary force field is essential here. And note potential energy increases at a given distance if when increasing the mass of the object/body, but is does not change if we deform the body (and change the pore) while preserving the total mass.

The energy form which relates to the matric potential has totally different invariances. The matric potential energy (I never came across this term) needs contact of a fluid with a solid surface. It would not vanish if gravity were switched off. It would however change when we mix the fluid with soap (at a constant mass), it would change if we deform the soil and change the pore size distribution (without changing the stored water mass) and it changes with temperature. Last but not least it does not increase but it decreases when we add mass to the system. This is why I think that it is not correct to treat it as potential energy - it clearly depends on the inner state and inner architecture of the system, so it belongs to inner energy! I regret when this statement is against the usual terminology of soil physics, but I think it makes much sense.

One option is to relate the matric potential to chemical energy, as done in a couple of studies cited above. This implies that the product of matric potential and gravitational acceleration g is then regarded as a chemical potential. We then end up with:

$dE=\mu dM + gzdM = (\rho g\psi d\theta + \rho gzd\theta)*V \rightarrow de= \rho g\psi d\theta + \rho gzd\theta$ (with $e=E/V$).

The point which is unclear is how to define the total chemical energy within a finite volume of the soil. At least Axel Kleidon and me were undecided about that. His idea was to integrate $de_{cap}= \rho g\psi d\theta$ from the residual water to the actual soil water content! This is consistent with the idea of a chemical potential, however, this implies that matric potential energy will increase with increasing soil water content and peak at saturation (note this is a definite integral). This is however, not what we measure and observe with the matric potential.

The other option is to postulate that we are in a similar comfortable situation as for potential energy. Here we can not only characterize it's change in but the total content of potential energy by integrating $\rho gz\theta$ the over the volume of interest (as for instance shown by Gerrit in one of his papers). This implies that

$\partial e_{free}/\partial V = \rho g\psi\theta + \rho gz\theta$.

and that we have to integrate the right hand side over the volume of interest to obtain the total energy of the water stored in this volume. I agree with Gerrit that we could easily start with this expression and proceed.

Yet is remains unclear what the baby is and how to name it. This is the reason for our effort to relate it to surface energy (which is in fact close to what Bolt and Frissel did). The rest is known from the manuscript. I agree that the assumption of cylindrical pores is a strong simplification, though it provides a link between matric potential and surface tension and thus surface energy. The reasons why we referred to Iwata is mainly because of the need to relate the change in surface area to a change in stored and thus of soil water content.

After a look to the avenue proposed by Bott and Frissel I agree that a start from Gibbs free energy has quite an advantage. They included the surface energy term into the VdP in their micro approach (their Eq. 6.5). This avoids the assumption of a cylindrical pore. (Yet also they treat what we call potential energy as separate term.) But their expression is dEpot= MGdh, which would corresponds in our terminology dEpot= $\rho g\theta V dz$), is much better suited for our argumentation and simplifies the entire derivation.

We will definitely revise the theoretical background section in line with Bolt and Frissel. I also agree that a more detailed discussion of the approach in comparison with your work is very much appropriate. But please note that we do not average. In fact we stratify observations along HAND difference and we account for variability among those, we can account for variability of retention curves as well (if known). The issue of the paper is also not about effective models, but to explain differences between observations. We will better stress this from the beginning and the above figure 2, shows the relevance for discharge and Figure 3 and 4 show that distributed modeling provides similar findings.

Language issues and overstating the novelity of the work

I now realize much better that the opening statement with a reference to the recent opinion paper of Savenije and Hrachowitz (2017) and the statement that "substantial attention in catchment hydrology since the work of Kleidon and Schymanski (2008)" might appear a little strange, as the main authors of these papers co-author the present manuscript. In this respect I'd like to stress this is my fault. I wrote most of the introduction before I started the discussion with Hubert Savenije and Axel Kleidon about the approach presented. So while the tenor of these passages does certainly reflect my personal point of view, a revision of the passages is certainly appropriate.

As already stated in our response to reviewer 1 it was not intended to overstate the novelity of the work. In fact in a former version of the manuscript it added several times phrase like "as it is well known". A few of my co-authors argued thermodynamics is not well known to most of the catchment hydrologists. Nevertheless we will revise this part and add the missing discussion of Gerrit de Roojs work.

Power = Q * ΔP (Q flow, P pressure)

The issue of relaxation times and time scales

Of course the soil runs behind these disturbances, yet they propagate into larger depths (and this is dissipative as well). With respect to relaxation times scales Figure 5 compares the soil water content in spring with the corresponding free energy state of a site in the Wollefsbach, note that e_{free} gets negative when the soil water content drops below 0.364 m^3m^{-3} (dashed line).

The time to drop into the negative range of e_{free} is 3-5 days, while the switching back to the positive range at day 65 occurs within a few hours during a rainfall event. It is not too much of a surprise that the time scale for drying the soil is much slower, as it is controlled by ET and especially root water uptake and hydraulic conductivity decreases with soil water content. Wetting is much faster, also when the soil is dry. This is due to the presence of preferential flow paths, which accelerate recharge and relaxation from dry conditions and export of excess water. (We have considerable experimental evidence about those in the catchments of interest and their dominance during recharge and runoff events (Angermann et al. 2018, Jackisch et al. 2018)). Figure 5 corroborates that relaxation times are fast enough to allow for perturbations around the local equilibrium state. We will better explain that drying and optional switching to the C-Regime operates at much longer time scales than wetting (and it needs roots to establish this perturbation in soil) and that the optional switch into the C-regime occurs at the weekly and not the event time scale.

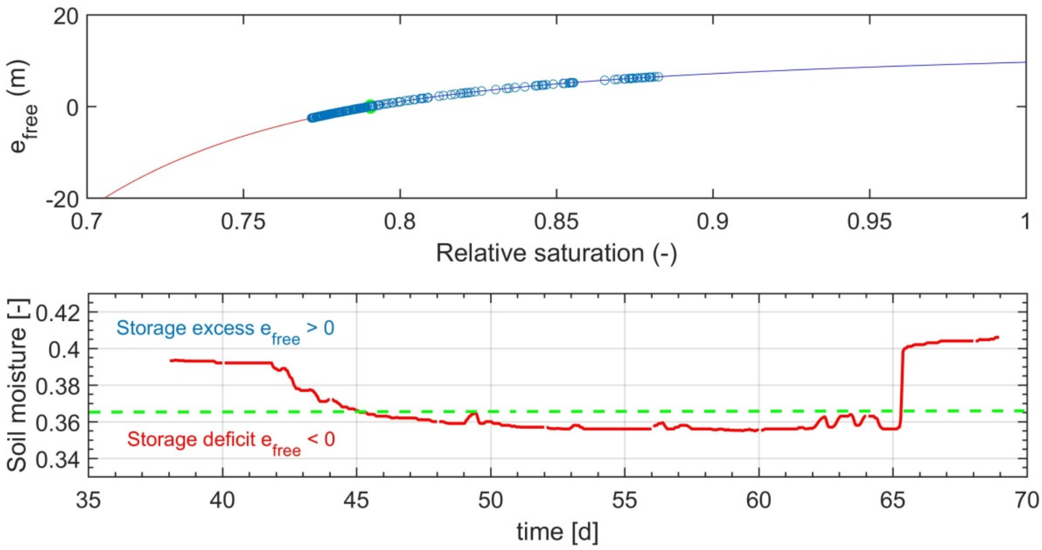


Figure 5: Free energy state and soil water content during a period of a month.

We have additional experimental evidence (joint observations of the soil water content and the matric potential) which corroborates that the energy state of soil water shows indeed perturbations around the energy minimum. This has been presented at EGU 2018 (Jackisch et al. 2018), which is currently on the way to a manuscript. And we have evidence that the approach allows indeed an reasonable estimation of depth to GW, based on known the retention function and the pair of matric potential and soil water content (as pointed out in the discussion of the manuscript). I see the point with the inertia, and I am very sure that aquifers operate and respond at much larger time scales. But I think there is no mechanical inertia in unsaturated soil water dynamics. The latter would imply kinetic energy or in other words an advection term in the Richards equation. But kinetic energy is as good as zero. I am also not aware of pressure that might propagate through the vadose zone (which requires inertia as well).

Casual use of water stock and thermodynamic equilibrium

I agree with Gerrit that we used the term water stock in a casual way, in fact we should use the soil moisture stock because this is the focus. I did not see that we doubt that the soil water stock depends on the all these influences highlighted by Gerrit (we will stress the role of vegetation as stated above). Yet if we want to characterize those we are thrown back to our observables: these are time series of matric potential, soil water content and if available depth to GW, and local retention properties. (We have no salinity measurements here (to infer on osmotic potentials) and I have no idea how to assess root water potential at all). If we combine these observables as suggested in our study, we learn much more about the temporal dynamics of the soil system at different locations in the catchment. This is not the entire picture but it is a way ahead (which I never saw elsewhere in this form).

I would say it the other way around, hydraulic equilibrium is a thermodynamic equilibrium, which corresponds to a state of minimum free energy/maximum entropy! The latter implies a zero potential gradient (including root water potential as well, we will state that we do not address this) in all directions.

In our study we do indeed focus on vertical distances, or more precisely HAND. I agree that this should be better stressed in the beginning. And yes we will be more precise in the sense that we look at the equilibrium of a distributed set of soil profiles spread along a HAND gradient. This is not the stock within the entire system, but it is a way to pool data, which are otherwise treated in a largely independent manner, into a joint sample (and explain differences) and to step beyond a mere look on comparing stored volumes!

Maybe we need to reformulate the story, as the proposed approach sheds light on whether the soil reaches the equilibrium or not. This is the case on the seasonal scale in the Wollefsbach (and the Weiherbach) but not in the Colpach.

<u>Calculations of water fluxes from averaged quantities</u>

I absolutely agree that you cannot calculate useful fluxes from averaged quantities without models and we will clarify this in the manuscript even stronger. I do not think that we ever claimed this, and in fact I think this is a key challenge to the REW approach. But as shown in the Figures presented above our approach can be used to inter-compare whether model errors are of gradual nature or whether the error is substantial (compare Figure 3), and to do this in a volume averaged manner as well.

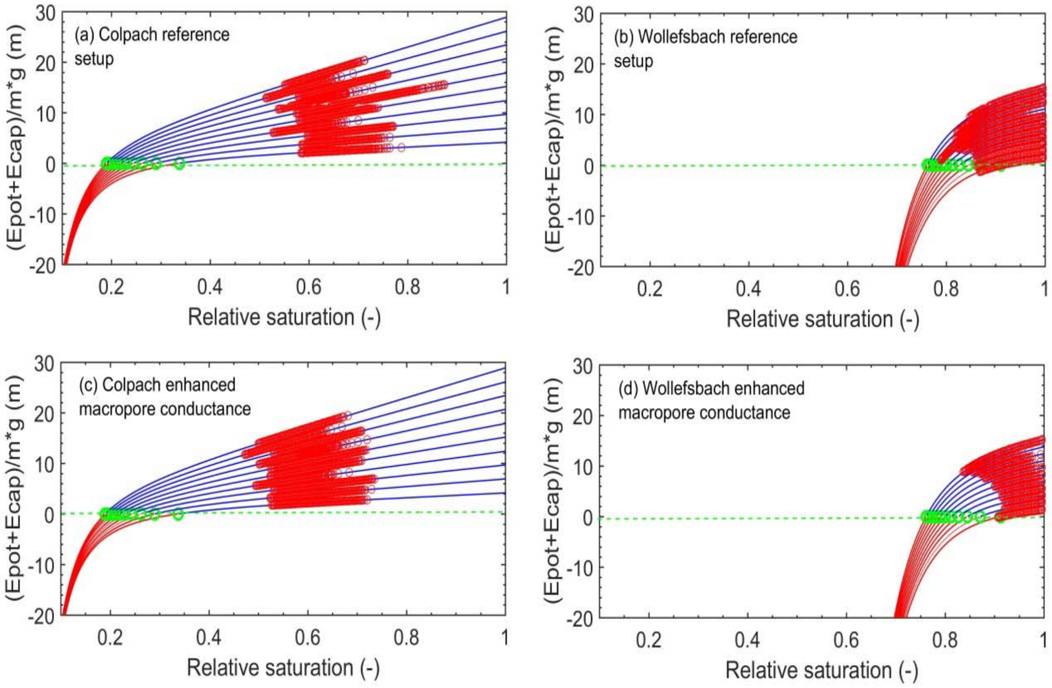


Figure 6: Sensitivity to changes in macropore conductance

And I think way to visualize soil water dynamics through its free energy state in the energy state functions /total potential functions is helpful to explore the sensitivity to changes in system characteristics. This is corroborated by Figure 6 were we compare the simulations from the reference model setups described in Loritz et al. (2017) with those where the conductance of the macropores in the system has been increased from 10^{-3} m/s to 10^{-2} m/s.

In both model setups the variability is reduced. The Colpach operates on average closer to the equilibrium states (due to an enhanced drainage efficiency) while the Wollefsbach operates at on average higher energy states, due to the enlarged recharge capacity.

Definition of the capillary surface energy?

See above, yes this fully relates to matric potential, but I stick to the point that this is not potential energy.

The energetic distance to equilibrium

This is an interesting point and in fact I started already to dig into Bolt and Frissel. And yes there is not much water needed to trigger strong changes in the energy state of the soil, this depends on the slope of the retention curve. Yet I think that our approach is valid, because it helps to balance this effect against the straight forward linear increase of potential energy (where the slope increases with HAND). There are breakeven points where either the one or the other dominates either when increasing saturation (this is well known) or when increasing HAND at constant saturation. It is novel to combine these two in a single characteristic. And this breakeven point corresponds to the threshold to trigger runoff response (if we focus on the riparian zone, compare Figure 2).

And yes the architecture of the pore space cannot be ingnored – so in case we go for volumetric averaging I would (for whatever purpose) I would rely on a calibrated model, which can account for this. The approach we propose can in fact account for variability in retention curves, if those were measured at the locations of the soil water sensors – but this is only partly the case.

Figure 1

Correct, I guess we remove it anyway.

Preferential flow and wetting (L 118 -119)

Yes I think that preferential flow for instance through a worm burrow, a plant root, or through does facilitate the recharge of dry soils. And there is plenty of evidence for that, for instance through reduction of overland flow production in Hortonian landscape if cracks emerge (Zehe et al. 2007) or macropore density increases (Zehe et al. 2005) or by deep routing vegetation in savannahs (Tietjen et al. 2009). If the flow path ends within the unsaturated zone, there is infiltration from the macropore into the surrounding soil. This contributes to wetting from the side (Beven and Clarke, 1986; Weiler and Naef, 2003; Klaus et al., 2013).

The derivation of our expression of free energy of soil water (lines 151 – 187)

Yes I think e_{cap} is not a potential energy- note that Bolt and Frissel in their eq. 4.1 account for potential energy as separate term as well.

In a previous version of the manuscript we started with equation 6. While I thought the link between matric potential and energy is straightforward, my co-authors disagreed and recommended a theory section to help those readers without a thermodynamic background through the argumentation. I still think that such a section is needed, but I agree the avenue of Bolt and Frissel is superior, because it avoids the assumption of cylindrical pores and includes the wetting angle.

I do not think that we recommended to forget about the matric potential. It is the other way around we recommend collection of joint data sets (matric potential and soil water content) as the product of water content and total potential provides more information about the state of the soil (and it is additive and can thus be averaged to larger spatial extents). Last not least the energetic view is also interesting opportunities to fit retention curves, as will be shown by Conrad Jackisch in a forth coming manuscript.

And I do not like the idea of a Newtonian analysis – simply as this comes from classical mechanics and thus essentially conservative systems. Friction and dissipation comes somehow through the back door. This may create a lot of misunderstanding (you may consult the argument between Ciaran Harman and myself in the discussion section of Zehe et al (2013)). Catchments and soils are open dissipative systems – the soil is so highly dissipative that soil water flows have almost no kinetic energy. Hence thermodynamics provides the right framework as dissipation is at its core.

Non-linearity and upscaling

Yes let's be precise with the issue of linearity: It is of course not new, but in my community often forgotten. The Darcy law has two sources of nonlinearity, one in the gradient term and the other in the hydraulic conductivity term (though I prefer flow resistance). In our study we just refer to the second and show that there is an breakeven point at which the gradient is dominated by gravity, which implies that free energy scales linearly with local soil water content. We could of course get this out by looking at the total potential as well. But not so nice instructive and potentials are not additive. Note that this dominance of the linear part emerges at sufficient high values of HAND, we will stress this. I am sure that Gerrit is true with his thoughts on the capillary fringe, but in fact we move to higher Hand distances here. So I do not think that this contradicts our findings.

And let's be precise with scales as well. By moving up scale we meant one hand that we increase the spatial extent of the domain by including more observation sites. Note that this does not include volume averaging, but that we stratify the observations along lines of similar geo-potential and saturation and plot them into the respective thermodynamic potential we call energy state function.

Of course we also included the point of volume averaging (which works see above) and this has in fact already been done in Zehe et al. (2013). Yet I prefer this stratified view against the average.

Section 2.3 and energy state curves are non-informative

I disagree: these curves provide much information about different drivers on soil water dynamics and how they change with HAND and with saturation. And they are very useful for inter-comparing soil moisture observations in a meaningful stratified manner, and distinction between the ranges of efree > 0 and efree < 0 is helpful to infer on the threshold for runoff generation (Figure 2). I never saw a graph that connects topography and retention is a meaningful and stratified manner. Many things I saw rely on averaging and I doubt whether this has any value.

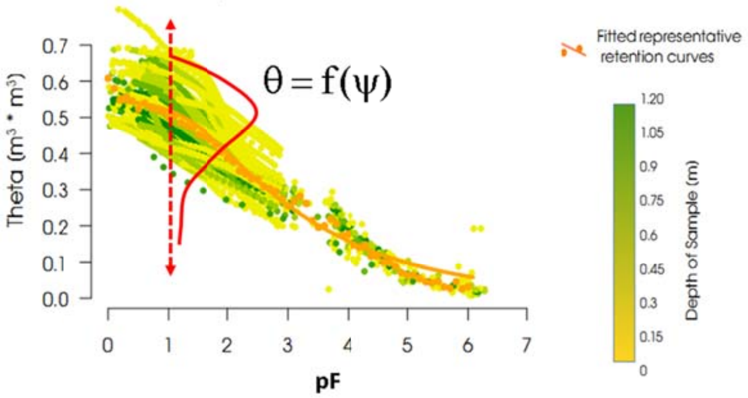
I agree with Gerrit that a corresponding plot for representative intervals of HAND might be better and we try this out (maybe with different line sizes to represent their fraction).

Do this perturbations across the equilibrium really occur?

They do, (see above) but not in all systems (not in the Colpach, not in Malallcahuello) and not on the event but on the weekly and seasonal scale. We will stress in the revised manuscript.

The representative retention function

I stick to this point and we additional have evidence for that. It is well known that retention in a fist size soil sample is not representative for heterogeneous systems (as the REV is much larger). So to me a set of experiments performed at 63 distributed soil cores yields a statistical distribution: As we control tension in these experiments, we observe frequency distribution of h (θ) conditioned by the tension. It makes absolutely not much sense to fit relations to single experiments and average the parameter afterwards (as it is done e.g. when estimating pedo-transferfunctions). The average parameter set will not represent the average relation between soil water content and tension. It makes sense to create a representative data set, and here we do this by averaging across all water contents of the experiments at a given tension (as given below).



$$\frac{\bar{\theta}-\theta_r}{\theta_s-\theta_r}=\frac{1}{\left[1+(\alpha\psi_S)^n\right]^{1-1/n}}$$

Figure 7a: Effective retention function of the Colpach

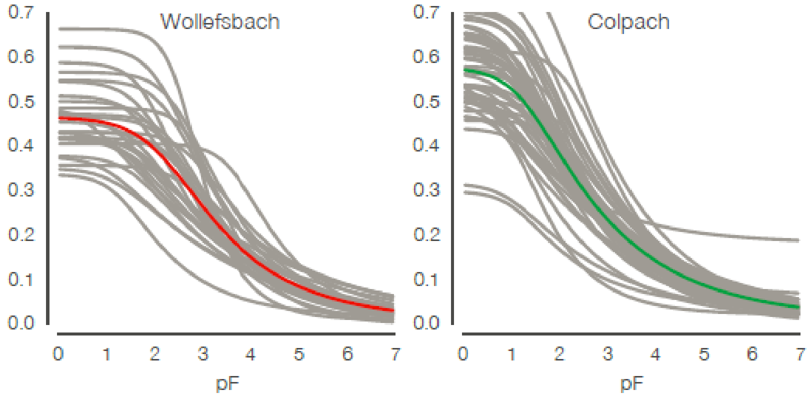


Figure 7b: Retention functions of obtained for single experiments

This corresponding retention function performs well in a simulation of the water balance of the Colpach using CATFLOW (find below the plot of total specific runoff versus specific discharge at the catchment outlet).

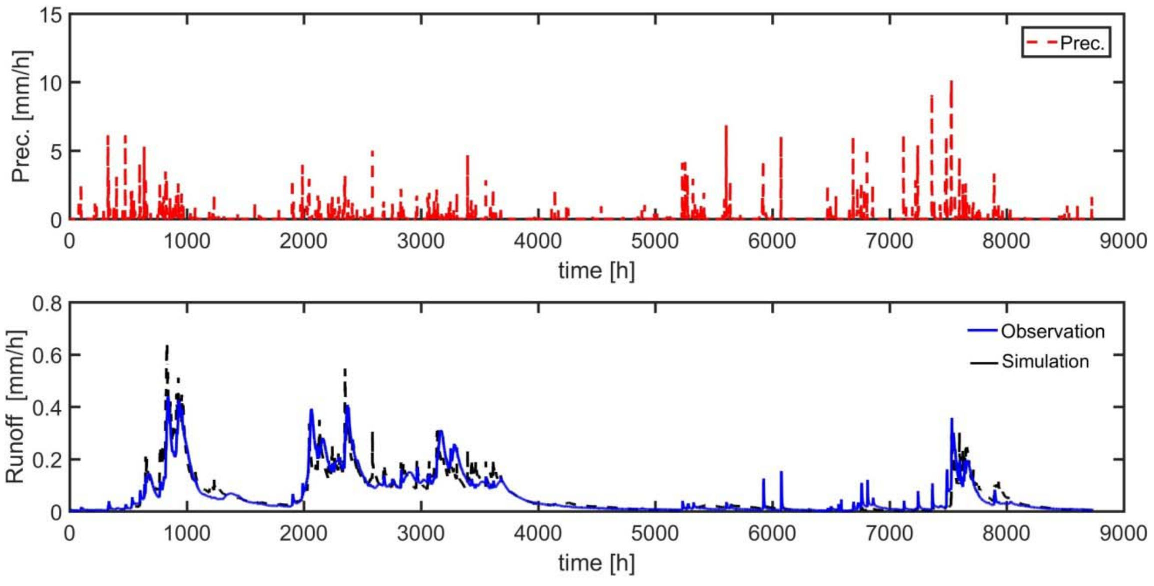


Figure 8: Simulated total specific runoff versus specific discharge at the catchment outlet (lower panel) and precipitation (upper panel) in the hydrological year 2013/14.

Please note that simulated runoff is strongly controlled by the wetness of local pools in the bedrock (both in the model and in the Colpach). Hence, simulated runoff is highly sensitive to the choice of the retention function as shown by the Figure 9.

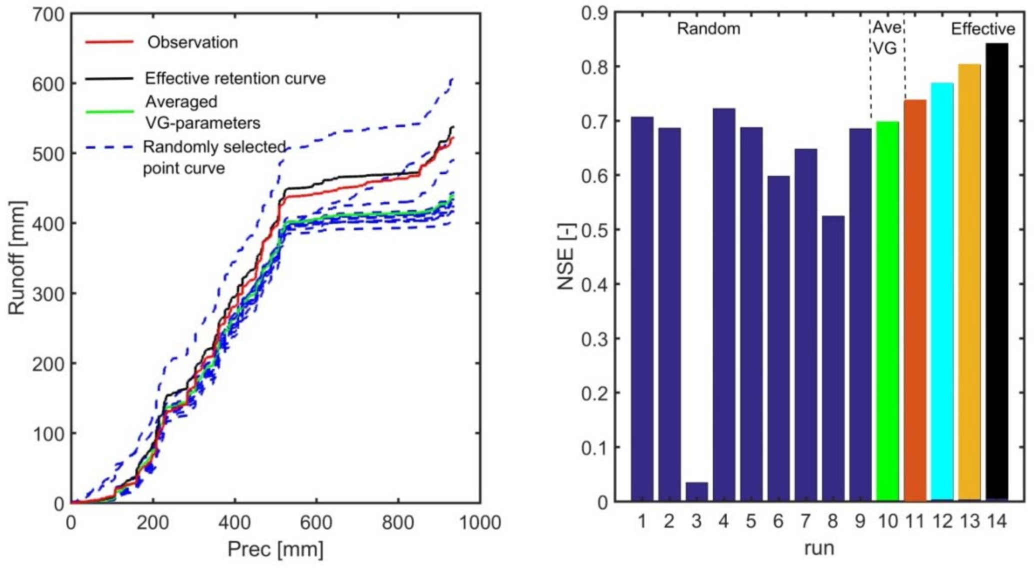


Figure 9: Sensitivity of double mass curve to the choice of the retention function (accumulated runoff plotted against accumulated precipitation in the hydrological year 13/14). The light panel shows the NSE for the randomly selected point curves (blue), the averaged van Genuchten parameters (green), the best model setup including the effective retention curve, the same configuration without vertical and lateral macropores. Note that this is unpublished work)

Yes observations is (matric potential + HAND) * theta.

Thank you very much for the stimulating discussion and the helpful comments

Erwin Zehe

References not given in the manuscript:

BEVEN, K. J., and CLARKE, R. T.: On the variation of infiltration into a homogeneous soil matrix containing a population of macropores, Water Resources Research, 22, 383-388, 1986.

Honerkamp, Josef und Römer, Hartmann. Klassische Theoretische Physik, Springer.

Conrad Jackisch, Axel Kleidon, Ralf Loritz, and Erwin Zehe; A thermodynamic interpretation of soil water retention and dynamics. Oral presentation at EGU General Assembly 2018. Geophysical Research Abstracts Vol. 20, EGU2018-12442, 2018

Klaus, J., Zehe, E., Elsner, M., Kulls, C., and McDonnell, J. J.: Macropore flow of old water revisited: Experimental insights from a tile-drained hillslope, Hydrology And Earth System Sciences, 17, 103-118, 10.5194/hess-17-103-2013, 2013.

Weiler, M., and Naef, F.: Simulating surface and subsurface initiation of macropore flow, Journal of Hydrology, 273, 139-154, 2003.