

Interactive comment on “Opinion paper: How to make our models more physically-based” by H. H. G. Savenije and M. Hrachowitz

H. H. G. Savenije and M. Hrachowitz

h.h.g.savenije@tudelft.nl

Received and published: 18 October 2016

We would like to thank the reviewer for his/her critical comments, which gave us quite a bit to think about. The comments highlighted that we need to clarify several crucial points as we do not seem to have managed to get our message across. Re-reading the manuscript it does indeed sound like a plea for “conceptual” models as preferred tools over “physical” models. That was not our intention as it is our clear understanding that (at least in hydrology) no model is completely physical and that all models remain to some extent conceptual. Rather, we wanted to emphasize the value of thinking at the macroscale and the resulting understanding of emergent processes as a stepping stone towards developing models that are more based on Darwinian thinking than the generation of currently available models. We have the impression that the reviewer did not fully realise that this paper is an opinion paper, with the aim to start a discussion

C1

on how we can make our models more physically-based, and that it is not a paper that wants to prove a point. The request for proof of concept is therefore not appropriate in this case.

Reviewer comment: In this paper, the authors claim that the current models are not physically based unless the model contains the characteristics of “an active organising agent”, which is the ecosystem. They therefore rely on the Darwinian theory which states that systems try to optimize themselves such that their potential for survival increases. This is indeed true for all living organisms in an ecosystem, which keep on evolving due to the interaction between both the organisms themselves as well as between the organisms and the surrounding abiotic environment, where latter may be changing, albeit through climate change or human impact as land use change or use of pesticides, . . . However, models do not work this way. Models are human inventions trying to describe nature in a formal language (dictated by mathematics), but remain inevitably with shortcomings as nature, due to its stochastic nature, cannot be captured in a few formulas.

Reply: We agree that models are indeed human constructs. However, we think that at the fundamental level they are rather dictated by our perception of the system, which we only then, as a second step, need to formalize in an as adequate as possible way using mathematics. It is also true that the current generation of models does not work that way. We think that this is a good reason to try and change that – why would we choose to work with models whose fundamental layout (i.e. time-invariant boundary conditions) does not reflect dynamic reality? Even in the presence of a wide variety of sources for uncertainties (observations, model structure, parameterization, parameters as well as stochastic fluctuations in the natural system) we think it is inconceivable to not even try to identify parts of the system that are and can be formulated in a dynamic way by making them a function of e.g. time (phenology affecting canopy interception throughout the year), vegetation (root zone storage capacity), wetness (soil cracks). Given the absence of detailed observations at the small scale, we think that

C2

understanding the processes and pattern that emerge at larger scales (landscape unit, catchment,...), is a valuable and feasible way to identify essential response patterns and their time-dynamics, which may then allow us to encapsulate this understanding in how we formulate our models.

Reviewer comment: Increasing the knowledge on processes in the model makes the model more complicated and often leads to numerous parameters that cannot be quantified through observations and require calibration. Furthermore, some equations used are extrapolated from the lab-scale to a model-scale of several meters to sometimes tens of kilometres, for which one could doubt whether the same type of equation remains valid if pixel-averaged values are used, yet, many papers can be found in literature where these equations (e.g. the Darcy equation) seem to behave well as the model outcomes seem to match the observations quite well. Taking an equation developed at the lab-scale to the grid-scale could be considered as a way of conceptualisation. Therefore, its parameters should therefore be interpreted with great care, even though the model is called 'physicallybased'.

Reply: Yes, we fully agree and do not state otherwise in our manuscript. This comment even corroborates one of our main arguments: neither most of our "physical" equations nor most of our direct observations can be meaningfully applied at the modelling scale of interest. This is not only true for lumped "conceptual" models but clearly also for distributed "physical" models, as these integrate over a certain grid scale often larger than 50-100m. Yet, equations were developed at smaller scales and parameters are most commonly obtained from (typically scarce or unavailable) observations at the point or plot scale. It is therefore not clear how both can be scaled up to adequately reflect the system including its sub-grid heterogeneity in a model. Here the large advantage of taking a step back and zooming out comes to our assistance: complex systems are very often characterized by process dynamics that only emerge as the scale of interest increases, reflecting the integrated small-scale processes and their interactions within the domain of interest. This relates back to the fundamentals of statistical physics

C3

(e.g. kinetic theory) and their application. As an analogy, consider the gas laws: the individual velocities and trajectories of molecules within a gas volume do not have to be known to understand and describe the system at the macroscale. Rather, a functional relationship between energy input and temperature emerges at that scale. It is therefore, in our understanding, not only possible but actually very likely that similar organizing principles characterize the hydrological response, as has been theoretically hypothesized (e.g. the work of Dooge, Sivapalan, Bloeschl, McDonnell, etc.) but also supported experimentally. For example, consider hillslope experiments that suggest that storm flow is generated at different moments at different parts of a hillslope, but that this distribution of connectivity thresholds emerges as rather stable functional relationship at the hillslope scale between hillslope-integrated storage and hillslope runoff (e.g. work of McGuire, Penna, Van Meerveld, McDonnell, etc.).

Reviewer comment: The fact ecosystems try to optimize themselves, in a Darwinian way, is, according to me, not a reason to state the models should be wrong. It is true that, if the system is changing rapidly, then the model's parameters may have to be changed in order to mimic the ecosystem's behaviour, but that's merely a calibration exercise that should be performed as soon as the model predictions (mean, variance, . . . or extreme behaviour) start to deviate from those of the observations. Such exercise should be done, irrespective of whether one uses a physically-based model, a conceptual model (which is advocated in the paper) or a purely empirical model. Any modeller should be aware that modelling a non-stationary system with a stationary system will result in a failure of the model if its parameters do not get updated (given that all hydrological processes are represented in the model).

Reply: We politely but firmly disagree. If a model cannot describe changing processes than it is, at the very least, quite incomplete. We are in fact rather surprised by the attitude that fully capturing the functional traits of a system should not be part of a model. Why would we want to re-calibrate a model if we could express the dynamic character of a changing system sub-component as a function of the driver that is controlling the

C4

change? Yet, to do so, we need to understand what is controlling the change. While this is, at the small scale, a task that is due to its complexity essentially impossible with current day observation technology, searching for larger scale emerging patterns that arise through organization of a system that keeps on ever adjusting and evolving is essentially the only possible way to proceed at this point.

Of course it is true that any modeller should be aware of the time-dynamic character of the system. Just as any modeller should engage in detailed model calibration/testing, such as multivariate model evaluation. Yet, browsing through literature, very few studies do either. Although parameter updating/re-calibration may be somewhat more common in operational forecasting, most publish literature uses models that calibrate over one single time period (and then often even omits any type of validation/post-calibration evaluation). More recently, with the increasing availability of long-term data, there is even a strong tendency towards calibrating using very long time-series (i.e. 40 years and longer), thereby essentially averaging out any type of variation, and resulting in a seemingly time-invariant system. This is of course not the case.

Reviewer comment: Conceptualizing physical properties, as presented in section 3 of this paper, is not solving this issue. The parameters of the individual stores need to be updated when the ecosystem changes. The paper does not at all address this, but presents the conceptual model as a solution for modelling non-stationary data.

Reply: Again, we thank the reviewer for this comment, but we also disagree here. Parameters only need to be updated if they are not formulated in a dynamic way. However, to formulate them in a time-dynamic way, we need to understand what controls the changes in the parameters, so that we can design a functional relationship (which, in the first instance may well be empirical). In our opinion, both examples in the manuscript, the landscape organization as well as the climatic controls on plant available root zone storage capacity, are steps into the right direction. Posing that different landscape types fulfil different hydrological functions (e.g. experimental work of McGlynn, Seibert, McGuire, etc) and that they therefore partition water in different ways,

C5

which is indeed not new as already Topmodel and many other models did that (yet often based on detailed soil information that is often not available), allows to design a catchment-scale model as the combination of functionally different landscape units (e.g. grassland and forest). After calibration each landscape unit will be characterized by different parameter values (e.g. grassland will have lower interception and root zone storage capacities). In case of (natural or anthropogenic) land cover change, e.g. deforestation, no re-calibration will be needed, assuming that the change in response dynamics is essentially controlled by the changes in interception and root zone storage capacity. The same is, to some extent, true for transfer of models to catchments they have not been calibrated for (Gao et al., 2016, WRR). The catchment-scale model therefore is a function of the relative areal proportions of the individual landscape classes used, which can be readily adjusted. Similarly, as there is very strong evidence for the role of precipitation P and evaporation E on the root zone storage capacity (S_{umax} ; e.g. Gao et al., 2014, GRL; DeBoer-Euser et al., 2016, WRR; Nijzink et al., 2016, HESSD), this parameter can be robustly estimated, depending on the data available, on e.g. the landscape unit or catchment scale. As the estimated values of S_{umax} are obtained from a functional relationship of P, E and S_{umax} , (1) S_{umax} does not have to be calibrated but can be directly used in the model and (2) S_{umax} in the model can be made a direct function of P and E. Any long-term change in P and E dynamics will then eventually give rise to changes in S_{umax} . Please note, that we currently do have a manuscript in review in HESSD, relating S_{umax} to landuse changes and a proof-of-concept to formulate S_{umax} as a function of time (in that case of forest regrowth; Nijzink et al., 2016). Again, this is practically feasible at the macroscale, therefore we gave emphasis to "conceptual" models in the manuscript, because they can deal with such adjustments rather simply. We will try to clarify that in the manuscript to avoid such misunderstandings.

Reviewer comment: In case the ecosystem is more or less in equilibrium, and thus stationary, then the problem of all types of models (physically-based, conceptual and empirical) becomes one of calibration, where e.g. root zone storage can be different

C6

depending on the plant species (several models take this into account: e.g. in CLM, only the moisture in the top layer is available to small vegetation, while for other vegetation types, the top two or top three soil layers (reaching more than a meter depth) are available. Other models (e.g. SWAP) take into account a rooting depth, root depth distribution or even root growth. In theory, one should be able to make these data pixel specific. Making them temporally variable is theoretically possible as well, but both (theoretical) options make the modelling exercise infeasible. Not taking this into account leads to (marginally?) larger uncertainties in the model predictions.

Reply: We agree, in principle, and as mentioned, the need for calibration is a considerable source of uncertainty. Epistemic observation errors and equifinality lead to the situation in which we fit wrong models, forced with wrong input data to wrong output data, inevitably obtaining parameters that do not represent the real world system in a meaningful way (e.g. Kirchner et al., 2006; Andreassian et al., 2009). But should it not be the main objective of hydrologists to describe the system in a way that it reflects reality and not (wrong) data as good as possible (even at the price of reduced model performance)? Thus, by direct observations at the modelling scale of interest (e.g. grid scale, catchment scale,...) these parameters do not have to be calibrated anymore (give or take observational uncertainties), thereby assigning a real physical meaning to the respective model component. In other words, if we use directly observed parameters (e.g. S_{umax}), we know that the real system operates like that. If this is reflected in models, no matter at which modelling scale or level of distribution, at least that part of the model is a direct reflection of real world dynamics. This in turn reduces the potential errors in other parts of the model. S_{umax} is an example of how this can actually be done at the macroscale. Of course, models can incorporate detailed root characteristics, etc. The relevant question here is, in spite of being theoretically correct, whether sufficient data are available at the required level of detail to actually provide a realistic representation in the model? We think that the answer is a clear "no" and that here, again, the zooming out can help to implement processes that can be actually observed (yet at larger scales) into a model, to make the model a better

C7

representation of real-world system functioning.

Reviewer comment: The plea made in the paper of taking ecosystem evolution into account in the models is a correct one, the solution provided, and the advocacy of a 'Darwinian approach to hydrology' is not convincing.

Reply: We agree that the manuscript would benefit from several clarifications to communicate our message in a way to avoid misinterpretations by the reader. We will give this point considerable attention to meaningfully address it in the revision.

Reviewer comment: While reading through the paper, I was very surprised, and disappointed, by: 1. The lengthy abstract including references. The abstract looks like a part of the introduction, and does not summarize what the paper is about.

Reply: We agree and we will redesign the abstract in the revision

Reviewer comment: 2. The language that is used. It is very informal, very often purely spoken language ("But there is more to it" (page 5, line 7); "No need to try and describe the sub-surface partitioning zone, . . ." page 10, lines 13-14; "This is good news for prediction in ungauged basins" (page 14, line 18). "And what is wrong with empirical formulations" (page 3, line 17), ". . ., but what many "physically based" models do not see" (page 9, line 13) – true models do not see, but one shouldn't phrase it this way –, . . .

Reply: Although this is an opinion paper, in which we think a somewhat more informal language is permitted to convey the subjective character of an opinion, we agree with the reviewer that some statements will benefit from being rephrased. We will do this in the revision.

Reviewer comment: 3. Sometimes, the text read as if it was the written down version of a presentation, where such statements are made, where statements may be made too bluntly, just to make a point. But, to me, such language cannot be used in a paper. Some examples where the authors make very blunt statements: (a) "hydrological

C8

models do an astonishingly poor job in . . ." (page 2, line 14). If so, then please prove this by demonstrating that no model ever could make predictions that were better than XXX (being an RMSE, NSE, . . .), where the XXX is a poor result. (b) "This inevitably results in models begin inadequate representations of real-world systems, haunted by frequently unreasonable model and/or parameter uncertainties and thus unreliable predictions" (page 2, lines 16-18). What does this mean? That the majority of our model is completely off? That we cannot rely any of our model results as a tool in water management? I believe we definitely can make use of these models, and many warning systems are adequate because they rely on model results.

Reply: This being an opinion paper, the above statements are pointed and exaggerated to emphasize that our models are far from perfect. We agree that our models are "not completely off" and that they are very useful for, in particular short term, predictions. Yet, what if climate variability changes as compared to the calibration period? What if other changes occur? Would it not be desirable if our models could account for that? At this point many models more often than not exhibit performance reductions when used in post-calibration prediction and/or they are characterized by an incapability to simultaneously reproduce multivariate performance metrics, indicating that they do not reflect the system internal process in a consistent way. This is linked to well-known error sources such as observational errors, unsuitable parameterizations and unsuitable parameter values and has been discussed and demonstrated by many authors (e.g. Klemes, 1986; Kirchner, 2006; Fenicia et al., 2008; Gupta et al., 2008; Clark et al., 2011, 2015; Andreassian et al., 2012; Euser et al., 2013; Hrachowitz et al., 2014; Zehe et al., 2014). We can of course cite a few studies as illustrative examples, however please note that this is not a review paper and we do not think that a detailed discussion on the literature is within the scope of this manuscript.

Reviewer comment: 4. The fact that the authors try to convince the reader, but, at least for me, completely fail in trying to explain what is wrong with our models? Going through Darwinian theory, which the authors try to apply it to hydrological models, is not

C9

at all convincing. The arguments that are brought are very poor (to me, blood running in veins is not an example that is similar to the water flowing in a landscape, I don't see the similarity, and that's only one example).

Reply: Pattern increasingly appear with increasing scale. They are a reflection of organization and a manifestation of the past evolution of the system. Organization does not appear from one day to the other nor is it a process that is finished. Rather, it is a continuously ongoing process, by catchments adjusting to ever changing environmental conditions in an efficient way. For example the root zone storage capacity S_{umax} is a product of the past climatic conditions and the vegetation that managed to adapt best to these conditions. However, the current day situation is not the endpoint and vegetation and thus root systems keep on evolving. The root system is at any time the result of vegetation that managed to find the most efficient trade-off between resource investment in sub- and above-surface growth to satisfy the contrasting priorities of sufficient water and nutrient access on the one hand and above-surface growth in competition with other types of vegetation on the other hand. The root system, however, also shapes (i.e. organizes) the hydrology to allow for distinct functions: generating storage to have sufficient access to water and efficient drainage of excess water (e.g. through preferential flow paths) to provide sufficient aeration for roots, such as our bodies evolved to organize blood flow through blood vessels, thereby permitting an efficient and simultaneous supply of oxygen to all our tissues. Our current generation of models are snapshots that describe the time-averaged status quo over the calibration period. Without further re-calibration, these models do not have the capability to meaningfully reflect ongoing and future changes to and ongoing organization in the system and are thereby incomplete representations of the system.

Reviewer comment: The fact that self-organization exists is true, but many models account for this, furthermore, some of these self-organizing processes are at time scales that are much larger than the average length of time series that are modelled (>100 year for the processes, to 10-50 year for time series, so, in the modelled time frame, the

C10

changes to the landscape may be very small, other impacts (such as land use change) will be much larger!).

Reply: We absolutely agree with that! This was also specifically stated with respect to preferential drainage features in the original manuscript ("at a wide range of spatial and temporal scales"; page 4, lines 21-23). While some processes change at the event scale (e.g. soil cracks), others change on seasonal scale (e.g. canopy interception), over scales of several years (root zone storage capacity) and even longer (e.g. soil formation, landscape formation, isostatic uplift, etc). Thus, some of the process will be relevant at time scales of hydrological predictions, other will be less relevant. These longer term changes will of course also be more problematic (or impossible) to identify with the available data (see also the work of e.g. Koutsoyiannis).

Reviewer comment: 5. In section 3, the authors try to explain why a conceptual model should and can do the job. However, this contains nothing novel, it is merely a description of a conceptualisation of physical processes as is commonly done in hydrology.

Reply: As discussed in the replies above, our main point here is that by zooming out to the macroscale (e.g. landscape class, catchment), processes that emerge from system internal organization at these larger scales can potentially be directly observed and used in the model. This allows not only assigning direct physical meaning to individual model sub-components at this scale, but these observations also allow the development of functional relationships between model parameters and their controlling factors, which can be the basis of a time-dynamic formulation of a model. We will try to makes this clearer in the revision.

Reviewer comment: 6. Stating that the fact that the baseflow demonstrates an exponential decay, and that groundwater is "organised" to flow to the river (page 13, lines 22-23), is, to me, not at all a fundamental question in hydrology. In fact, I cannot cite papers where this issue is being brought up as being a fundamental question.

Reply: It is indeed true, that there are not many papers that address this topic (e.g.

C11

Fenia et al., 2006). Yet, we think that understanding the reason "why" exponential recession is ubiquitous over a range of scales and how it is linked to Darcy, is a fundamental question that, if answered, could teach us a lot about how heterogeneity in hydrological systems integrates to larger scales and what controls this integration. A linear reservoir of course reminds of Darcy, when interpreting the recession constant k as product of effective (i.e. lumped) hydraulic conductivity. But the linear reservoir can only emerge if the drainage pattern towards the open water is in agreement with the shape of the groundwater water body feeding it. It requires a functional descriptor of catchment "form".

Reviewer comment: [...]the fact that the baseflow demonstrates] an exponential decay is merely caused by the boundary condition that a river poses to the groundwater body. Stating that it may have to be attributed to the theory of maximum entropy production seems inappropriate if not wrong. In fact, any organized system has a low entropy. Maximizing entropy should not lead to increasing organization of flow lines. . .

Reply: This seems to be a misunderstanding. Of course an organized system is characterized by low entropy. We do not state otherwise. Rather, we argue that hydrological systems seem to maximize entropy production, i.e. the rate of change of entropy (which of course can also occur at low levels of entropy). Seeing entropy in a general way as a gradient to be dissipated, the hydrological system has indeed low entropy, as it is characterized by a large gradient (i.e. hydraulic head as essentially imposed by topography). The idea is that the system may evolve towards a state that allows it to dissipate this gradient as efficiently as possible, i.e. by efficiently eroding the landscape until no more elevation gradient is present (which of course is counter balanced by dust deposition and uplift), but also providing an efficient surface drainage network that can transport the eroded material (see Kleidon et al., 2013) An interesting hypothesis would be that the linear reservoir behavior is a consequence of the most efficient depletion of the potential energy of the active groundwater in the catchment.. The river that poses the downstream boundary condition was not always there. Rather, following

C12

the work of Dietrich, Montgomery, Tarboton et al, it gradually developed through an interplay of convergent flow (organization!) and landscape gradients, which both provide the energy supply for forming channels through erosion. Thus, the exponential recession may very well be a manifestation of maximized entropy production. And why it is exponential, we do not know, but are curious to learn (see reply to comment above).

Reviewer comment: 7. Section 4 deals with the practical consequences. Actually, whatever is in this section has, to me, nothing to do with "Darwinian theory" but is merely (1) a listing of some advantages of conceptualization of hydrological processes. (2) a demonstration of the usefulness of DEM info. The fact that DEMs can be used for discerning between different hydrological processes is well known (actually one of the most famous hydrologic models, i.e. the Topmodel, is based on this). So, again, nothing new is stated here.

Reply: This relates very much to the replies given above. We tried to convey that zooming out to the macroscale permits us to actually observe processes (or their controls) on that scale and that these observations can be directly used in a model, representing real processes and that scale and not being mere calibration artefacts. And yes, Topmodel uses landscape information as well as do many earlier or later models, too, but it is often not seen in the context of organization, which would provide further advantages as detailed above. Regarding the use of Darwinian theory, the section reports on the derivation of $S_{u\max}$ as the result of evolution, whereby ecosystems adjust to the climatic drivers that determine their conditions for survival. Of course any hydrological model uses mass, momentum (although represented by closure relations) and energy conservation. But the presented approach uses evolutionary thinking to parameterise these relations by physical quantities (such as storage capacities and residence times or process time scales) that have direct physical meaning and can be derived directly from data without the need for model calibration. Although this approach can potentially be used in all hydrological models, conceptual models are, thanks to their mathematical simplicity, very fit for this purpose.

C13

Reviewer comment: (3) an alternative way of estimating a model parameter (in this case the root zone storage capacity). The fact that plant root depth depends on the climatological condition, is known as well. Many models indeed approximate the root zone depth using a constant value, and this is wrong, but, this is a simplification (amongst the many others) of the model. (Actually, all conceptual models are highly simplifying physics). Trying to estimate the root zone storage capacity from E and P, can be done as is stated, and this value can be used in a model, but, due to all simplifications, one might end up with a sub-optimal model if this value is used. . .

Reply: We disagree here as we think that at the macroscale, physics are not simplified as in "approximated" (as we think the reviewer intends to say), but they are simplified as in "simple patterns emerge at large scale", which makes them no less physical. On the contrary, these macroscale process actually provide us with means to understand the effect of heterogeneity (which is in the common absence of observations at suitable spatial resolutions not warranted by smaller scale representations). In other words, we think that the macroscale observation (or estimation) of a process is NOT a simplification. It is actually what we need to do (besides adapting our equations to that scale) if we want our models to capture natural heterogeneity. And yes, fixing a model parameter to a directly observed value will very likely reduce model performance. And we think that this is (under the given circumstances such as uncertainty in data and parametrization) extremely good news, as it forces the model to reproduce dynamics that are observed in reality, allowing to identify and discard parameterizations and other model parameters that merely provide good fits but that are mere "mathematical marionettes" (Kirchner, 2006). Thus, at the price of calibration performance, it makes our models better representations of the observed hydrological system.

Reviewer comment: Furthermore, what is the link between Darwinian theory and the calculation of root zone storage if the latter value is not "continuously" changing (as Darwinian theory considers systems to evolve to an optimum)? How should conceptual models take this into account? As far as I understand the paper, only once, E and P

C14

time series are used to fix the root zone storage capacity (not to evolve in time).

Reply: We think that before thinking of implementing a dynamic parameterization, we need to understand the process and functional relationships emerging at larger scales (thus fix $S_{u\max}$ to one observed value). Once we have identified and understood these, we can design a dynamic formulation. For $S_{u\max}$ we followed this sequence in a series of recent papers, the last of which actually provides a proof-of-concept for a time dynamic formulation of $S_{u\max}$ following deforestation and regrowth (Gao et al., 2014, GRL; De Boer-Euser et al., 2016, WRR; Nijzink et al., 2016, HESSD). The reviewer is correct in assuming that a time dynamic implementation requires a temporally changing $S_{u\max}$ as dictated by P and E time series.

Reviewer comment: (4) an alternative way for estimating the recession constant of the groundwater reservoir by using GRACE data. But one could argue that for very small catchments, the recession found by GRACE data is not to the same as is found for the recession curves in the hydrographs of the catchment due to scaling differences. Furthermore, the idea of using GRACE for estimating the recession parameter of a conceptual model is not verified. It is only bluntly stated that the recession of Sg obeys the same (?) exponential function as the recession of the drainage network (page 17, lines 12-14). Such statement at least requires a demonstration that the same (!) exponential function (or at least recession constant in this function) is indeed found.

Reply: The reviewer is right that there is no proof of concept yet on using GRACE for recession analysis. But in an opinion paper this is not needed. Rather it suggests a new approach that can become valuable as better gravity information will become available over time. The same can be said about the E and P products. We can safely assume that these products will become better by the year. Therefore, the approach sketched here is in our opinion an interesting way forward to explore. If we have given the impression that this is all established science, then we shall modify the paper so as to avoid this impression.

C15

Reviewer comment: (5) a plea for using remote sensing data as additional source of data. But what is "Darwinian" about this?

Reply: Indeed, there is not much Darwinian in remote sensing. Yet the remote sensing derived products of E and P allowed us to derive $S_{u\max}$ by an evolutionary theory. Moreover, the paper is not merely about Darwinian thinking, it is about making our models more physically-based. In order to do so, we want to use observable quantities that represent states and fluxes in our models. Remote sensing is a crucial toll for this, since it allows observations at the appropriate system scales, even covering parts of the world where ground observations are limited.

Reviewer comment: To a certain extent, I can fully agree with the authors that fully fledged physicallybased models are so complex to work with as they have too many, non-observable parameters and therefore too many parameters to calibrate that they do a poor job, while working with simple conceptual models that simplify the physical processes e.g. in a set of reservoirs, can provide much better results (at least if the model is used for predictions but not for trying to get detailed insight in the system: e.g. particle tracking cannot be done, but may be necessary if one wishes to figure out the source of a pollution).

Reply: If not particle tracking, than at the very least flux tracking is clearly possible using models at the macroscale ("conceptual"). There was quite some progress over the last years on this topic, with many models having been shown to have the skill to reproduce the dynamics of both conservative and non-conservative solutes together with the ability to track the actually solute and water movement through the system (e.g. Birkel et al., 2010, 2012, 2015; Botter et al., 2008, 2010; Rinaldo et al., 2011; Benettin et al., 2013, 2015; Hrachowitz et al., 2013, 2015, 2016; Queloz et al., 2013; Harman, 2015), the state-of-the-art being a theoretical framework of how catchments store and release water (Rinaldo et al., 2015), which can, if warranted by data also be implemented in a spatially distributed way.

C16

Reviewer comment: This is actually the main message of the paper, but by making use of theories that are not suited for it, in order to try to make this point. Furthermore, the applicability of these theories on hydrology is not demonstrated with specific examples, while many statements are made bluntly, mainly because too often, spoken language is used instead of a formal written language. Given the above comments, this paper seems the result of a wild and out-of-the-box thinking, without any proof or demonstration of its potential, and furthermore, the language used is inappropriate for a paper, I believe it cannot be accepted in this shape, but needs more work, making it a state-of-the-art scientific text (without the spoken language), including proofs of the statements that are made and making sure that the final sections (3 and 4) support the theories being argued in the first two sections. Finally, referencing to a poster, which is included as appendix, should not be done. Rather, the material in the poster, if relevant, should be included in the paper.

Reply: We think that the reviewer clearly does not realise that this is an opinion paper. Opinions don't require any "proof or demonstration of its potential" a sound reasoning should be enough to trigger debate and exchange of views. Judging from this reviewer's comment, this objective has certainly been met. Also the language should not be formal. An opinion should be clearly expressed without the use of formal or less clear language. It is clear that the opinion is ours. Using the first person makes that clear to everybody, which by using the impersonal passive form would not at all be clear. Maybe the paper is wild. But we certainly hope that it is out-of-the-box, which is what one would expect an opinion paper to be.

Selected references:

De Boer-Euser, T., H. K. McMillan, M. Hrachowitz, H. C. Winsemius, and H. H. G. Savenije, 2016. Influence of soil and climate on root zone storage capacity, *Water Resour. Res.*, 52, 2009–2024, doi:10.1002/2015WR018115.

Fenia, F., H. H. G. Savenije, P. Matgen, and L. Pfister, 2006. Is the groundwater

C17

reservoir linear? Learning from data in hydrological modelling. *Hydrology and Earth System Sciences*, Vol. 10: 139–150.

Gao, H., M. Hrachowitz, S.J. Schymanski, F. Fenicia, N. Sriwongsitanon, H.H.G. Savenije, 2014. Climate controls how ecosystems size the root zone storage capacity at catchment scale, *Geophysical Research Letters*, 41, 7916-7923, doi: 10.1002/2014GL061668.

Kleidon, A., Zehe, E., Ehret, U., and Scherer, U.: Thermodynamics, maximum power, and the dynamics of preferential river flow structures at the continental scale, *Hydrol. Earth Syst. Sci.*, 17, 225-251, doi:10.5194/hess-17-225-2013, 2013.

Nijzink, R., Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., Han, D., Wagener, W., McGuire, K., Savenije, H., and Hrachowitz, M.: The evolution of root zone moisture capacities after land use change: a step towards predictions under change?, *Hydrol. Earth Syst. Sci. Discuss.*, doi:10.5194/hess-2016-427, in review, 2016.

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, doi:10.5194/hess-2016-433, 2016.

C18