Reviewer #2 (Ralf Loritz)

Comment:

I do agree with the authors that the discussion about the different modelling philosophies is sometimes rather driven by emotions than by facts. I also think that an opinion on this issue and proposals for a way forward could be of interest for publication in HESS. However, I believe that before this paper can be accepted for publication substantial revisions are needed.

Reply:

We highly appreciate the reviewer's positive, open and highly constructive comments. He raised quite a few points that made us reflect more about the actually underlying issues and we will incorporate his comments as fully as possible in the revised version of the manuscript.

Comment:

First of all, both authors have a separate opinion paper or comment with a closely related content in HESSD at the moment (Clark et al., 2017; Savenije and Hrachowitz, 2016). Especially the discussion and the review of the opinion paper by Savenije and Hrachowitz (2016) cover a lot of similar points and arguments as this paper. But also the comment by Clark et al. (2017) has several overlapping arguments, especially related to the proposal about how to progress in hydrological modelling. With three papers in HESSD covering similar topics I think it is especially important that the authors clearly show what this opinion paper differentiates it from the other two manuscripts.

Reply:

The reviewer is right in pointing out that we have separate opinion papers, either in review or recently published in HESS. That was not planned and as often in life, things frequently seem to temporally culminate. As a background information: we, the authors, participated in last year's workshop on "Improving the Theoretical Underpinnings of Hydrologic Models" in Bertinoro, Italy. Among the other three dozen participants were some of the most experienced modellers in the discipline of hydrology. Notwithstanding this high level of expertise, one of the most (and most emotionally) discussed topics during this workshop was the difference between different modelling strategies as well as their respective theoretical/physical basis (and lack thereof). As we found these, and further discussions after the workshop among ourselves, the two authors, highly instructive we believe that sharing the different points of view and offering some sort of synthesis may help to direct future efforts in modelling towards more effective developments.

Quite naturally, the resulting manuscript then aimed to communicate our opinion on how we, as community, need to understand and approach the modelling problem, which touches the core expertise of both of us, and is thus somewhat related to our respective ongoing work. Notwithstanding the same general topic, i.e. the state-of-art and future needs for modelling, we actually do not see too much overlap between the mentioned papers. Here, our main objective is to <u>*resolve the perceived dichotomy*</u> *between different modelling strategies, which is, in our opinion exactly and exclusively that: perceived. In other words, we intended to make the point that all models* *are fundamentally the same and that they mostly only differ in their degree of resolution (i.e. complexity): what is the spatial resolution of the model domain (spatial complexity)? Similarly, to which degree do we resolve or lump different processes in our representations (process complexity; see example in S2)?*

In contrast, on the one hand the Savenije and Hrachowitz (2017) paper emphasizes the need to account for system characteristics that evolve over several spatial and temporal scales if we want to improve our understanding of the hydrological system but also our predictions. On the other hand, Clark et al. (2017) provide a general discussion of (amongst others) the need for a better understanding of scaling in hydrological systems, without making the direct link to top-down/bottom-up models.

In this sense, we believe that our manuscript provides additional value by providing a synthesis and suggesting a more stream-lined approach to modelling, arguing that the actual challenges lie in identifying parameters at the relevant scales and which equally apply to both (perceived) endpoints of the modelling spectrum.

In any case, we will provide a clearer distinction between the mentioned papers and provide a clearer positioning of this manuscript in the context of existing literature.

<u>Comment:</u>

My second concern is that a substantial part of this paper reads like a text book. While the language is clear and easy to follow, I was wondering if the potential audience really needs a two page long introduction to “conceptual” and “physically-based” models? Similarly, other sections seem to be redundant as they have already been covered in great detail in several opinion, comment and review papers (e.g. Bahremand, 2015; Clark et al., 2015; Gupta et al., 2012).

<u>Reply:</u>

We acknowledge this point raised by the reviewer. The reason we included a short background on different models was that we think many of the discussions around the use of a specific modelling strategy arise from miscommunications and misunderstandings. We agree, that a modeller will understand and interpret his/her model together with its advantages and disadvantages in a meaningful way. However, we also think that any other modeller will do so in a different way. For example, is there a clear understanding in the community that conceptual models originate from lumped unit hydrograph approaches and that what they essentially do is reproducing observed dispersion characteristics in a signal processing sense and that they can be implemented at any level of complexity (see above), which comes along with the need to converge towards physically based models? We do not think so. Otherwise it would be surprising why many modellers dismiss conceptual models per se as having no physical basis (which may be true for specific implementations, though). To avoid these types of misunderstanding, we believe that in a paper that intends to provide a synthesis of the situation, at very first <u>common ground needs to be established</u> to avoid further misunderstandings. Therefore we also think that a short description of different approaches needs to be part of this manuscript. However, we will change the section to provide an actual framework for a more rigorous model taxonomy.

We agree, that some points discussed in our paper have already been covered previously. However, in most cases only individual aspects were discussed. While for example, Bahremand (2016) emphasises the need for parameter allocation to replace calibration, Clark et al. (2016) put the focus on the value of synthesis of hydrological understanding for developing testable model hypotheses and the associated need for more rigorous model evaluation. In contrast, the main intention here is to develop and communicate the point that all model types have, if well implemented, a robust physical basis, albeit at different scales, and that they essentially share the same problems (e.g. need for calibration, hypotheses that are difficult to test with available data, etc.). We will make this clearer in the revised manuscript.

<u>Comment:</u>

This brings me to a more general comment aimed at all opinion papers which is that careful reading is required to identify where facts end and the opinion of the authors starts. One example for this paper is when the authors write that top-down models have "a parsimonious representation of the energy balance". Is this a fact and has it been shown somewhere or is this an opinion? As far as I know, most hydrological models do not close the energy balance or even keep track of the energy in the system. How can you know if you close the energy balance, when you only try to close the mass balance?

<u>Reply:</u>

This is an interesting point, which we are glad to clarify. We fully agree that most models do not track energy through the system in a detailed way due to the complexity of the processes involved and the lack of data to meaningfully constrain/test potential model formulations of these processes. However, this does not mean that energy is not considered.

Energy input is a first order control on the partitioning of water fluxes into drainage and evaporation/transpiration. This partitioning is (or better: needs to be) present in any model. Posing that potential evaporation is a meaningful proxy for incoming energy, the modelled actual evaporation/transpiration then approximately closes the energy balance if: (1) the modelled partitioning between drainage and evaporation/transpiration reflects the observed partitioning (i.e. runoff coefficient) and (2) there is negligible inter-catchment groundwater flow. While the latter point is arguably difficult to test in most catchments, a model can be constrained to reproduce a meaningful partitioning pattern by not exclusively calibrating it to stream flow, but simultaneously also to the runoff coefficient (e.g. long-term average, inter-annual and/or seasonal). This is quite evidently a simple black-box approach to the energy transfer in hydrological systems but it allows the system overall energy balance to be approximately closed. In other words, the energy balance is implicitly and in a simplified way present in the runoff coefficient (see also the Budyko relationship). From that perspective, if well implemented (as stressed on page 6,l.24-27 in the original manuscript), top-down models do at least not considerably violate the energy balance. Even if this is not explicit, we do not think that this is an opinion but rather a <u>physical necessity</u>. We will further clarify this in the manuscript.

Comment:

Another example is the unclear separation of the macro- and microscale in this paper. For instance the authors argue that macroscale models are important and physically-based with e.g. Sivapalan's (2005) search for a general law at the macroscale or with a comparison with Gay Lussac's law. However, the papers they mention to support this argument use often macroscale models to define various states of the microscale, for example the root zone storage. While using macroscale models to estimate states at the microscale is a perfectly valid approach, it is very important to make clear to the reader that this can only be an estimate and is rather difficult because of the high degrees of freedom we have in hydrology. A precise definition of the macro- and microscale and a clear structure of the manuscript in this context might help to improve this paper and would ensure that not even more "modelling myths" are generated.

Reply:

This is a very good point and we fully agree with the reviewer. We will provide a clearer definition of macro- and microscales. We understand the microscale as the scale at which direct observations of the system boundary conditions/parameters are typically available, i.e. soil sample, plot scale, individual plants, etc. These values emerge from yet smaller scale processes and heterogeneities at the scale of the actual observation and are fully valid for the domain they have been determined for. However, and quite obviously, they cannot by themselves account for spatial heterogeneities between the sampling points and the feedback effects arising from these. For example, interception capacity can be determined for some individual plants (or groups of plants; i.e. microscale) but is problematic to transfer to other parts of a system or to scale-up as it is influenced by a range of different factors, including but not limited to plant species, plant age, plant shape, plant location (wind exposure!), composition of different plant species, or spatial densities of individual plants.

The macroscale is then, in our understanding, the scale at which the heterogeneities of the individual microscales observations and in-between are integrated to emerge as functional relationships (see example S2). These are not directly observable using standard observation technology. Yet, there is potential for quite robustly inferring at least some of them through analysis of domain (e.g. catchment) scale data (e.g. runoff). Some examples include the time scale of the groundwater or the root zone storage capacities.

The reviewer is correct in assuming that macroscale values are then estimations and it is, as they represent the domain integrated picture, difficult to infer spatial patterns beyond the domain they have been developed for (e.g. catchment, HRU, grid cell, etc.). In other words, when using a lumped model with a lumped root zone storage capacity, this capacity will very well represent the system overall capacity, but intra-domain spatial differences cannot not be readily extracted. On the other hand, basing the root zone storage capacity on microscale values, i.e. point observations in a distributed, bottom-up model formulation will allow a representation of spatial pattern. This, however, comes at the price that the spatial heterogeneities between the (typically scarce) observation points and therefore the system overall capacities are likely to be not well captured, thus introducing uncertainty.

In this context we politely disagree with the reviewer, because we do not think that a macroscale parametrization does result in more degrees of freedom than a microscale parametrization. It is true that at this point most macroscale parameters cannot be observed and tested against data and therefore require calibration (with all its adverse effects). But the same is true for microscale parametrizations: they are strictly valid for their points of observations but not beyond that. As we do not have spatially seamless observations, using these parameters to describe the entire domain either provides a false sense of model accuracy or requires additional calibration (for a much higher degree of freedom than the macroscale representation). Thus, to be cheeky: there is no free lunch, or both ways currently still have substantial drawbacks.

<u>Comment:</u>

However, I believe that we do not need another paper where we discuss how physically-based or not the different modelling philosophies are. I recommend that you focus on the complementary merits of both approaches. Furthermore, I suggest giving clear examples and sharing your ideas how we could for instance combine top-down and bottom-up models in practice. This could make the manuscript much more unique and meaningful. As I believe that the discussion of this topic is of relevance for the hydrological community I hope my comments, questions and opinions are constructive and can help to improve this manuscript.

<u>*Reply:*</u>

We agree with the reviewer and we will give more emphasis on the complementary merits of the approaches and how these can best be exploited. We think that a fruitful way forward will be to let spatial and process complexities of conceptual models converge towards the representations in physically based models and vice-versa, with the overall aim of formulating models that can better satisfy the contrasting priorities of (1) a meaningful representation of in particular spatial patterns and (2) meaningful tests of the underlying hypotheses while (3) keeping the required degrees of freedom at a minimum level.

<u>Comment:</u>

Page 2 Line 28: Maybe add some references where the authors showed that their model failed after the calibration period, both from the bottom up and top down community.

<u>*Reply:*</u>

Agreed, we will provide some references.

<u>Comment:</u>

Page 3 Line 6: Could you define catchment scale?

<u>*Reply:*</u>

We define catchment scale here as all scales starting from a few hillslopes that are drained by a 1st-order stream.

<u>Comment:</u>

Page 3 Line 10-11: What do you mean here with "respect to bottom up models".

<u>Reply:</u>

We will re-formulate this sentence.

<u>Comment:</u>

Page 3 Line 11 – 12: I couldn't find the part where you provide a perspective of how to take advantage of different modelling philosophies.

<u>Reply:</u>

We will clarify that and put more emphasis on this aspect in the revised manuscript.

<u>Comment:</u>

Section 2 Modelling philosophy: This section is mostly written clearly and precisely. Nevertheless, I think the potential reader of this opinion paper is already familiar with the different modelling approaches and reading this section is very akin to reading a text book. I would consider shortening this section with references to other studies or textbooks.

<u>Reply:</u>

We fully agree that the reader will be familiar with the different approaches, but not necessarily with their origin/background. This and common misunderstandings do in our opinion call for the need to establish common ground (see reply to comment above). However, instead of providing the general backgrounds of the two model approaches, we will give a wider view and provide a more systematic framework for an actual model taxanomy.

<u>Comment:</u>

Page 4 Line 4 – 5: From my point of view the scenario in which you end up in a catchment where you only have reliable runoff and rainfall data but nothing more available is rather unrealistic: In which catchment in the world do you have reliable streamflow, evapotranspiration and rainfall measurement but no other information of the catchment? At least in Europe and the US you have land cover and geological maps. Furthermore, if there is a gauging station and rainfall measurements, most likely a person is doing maintenance on the respective instruments on a regular basis. This

person will most likely accumulate a lot of qualitative information about the hydrological functioning of the catchment and could possibly also complement this picture with low-effort additional measurements or soil sampling. For instance Jackisch et al. (2014) showed how fast one can characterize a remote meso-scale catchment based on a brief measurement campaign. If land cover is managed forest or agriculture, frequently nationwide reports on productivity and for example drought risks are available. We have digital elevation models for the whole earth in decent resolution, monthly estimates of precipitation and soil moisture from satellites and so on. In my opinion the problem is often very different from the projected scenario: We do not know how to use the data in our hydrological models or if it is of relevance. But I admit that this may be a different story.

Reply:

We fully agree with the reviewer and did not intend to imply otherwise. What was meant here is hydrometeorological data. As at this point no generally valid, reliable and quantifiable functional relationships between factors such as topography, geology or soil types on the hydrological response are available, these data are very valuable to develop and constrain models but cannot be a stand-alone replacement of actual hydrometeorological and hydrological observations (i.e. precipitation, stream flow, etc.). In the absence of detailed, spatial high-resolution observations of fluxes and states (cf. the boundary flux problem), conceptual models are therefore mostly developed on the basis of what is available, which is in most cases stream flow, precipitation and estimates of potential evaporation. Of course we agree with the reviewer that the mentioned system characteristics should and eventually need to be used to develop meaningful models, a point which we explicitly address in section 3.2 and its subsections.

Similarly, we fully agree that we need to be more efficient in extracting information from our data, which boils down to the paragraph starting at page 8, line 27 in the original manuscript: "The lack of an adequate model calibration, testing and evaluation culture partly arises both from insufficient exploitation of the information content of the available data, and also the real lack of suitable data to more effectively constrain models […]"

Comment:

Page 4 Line 5-8: Is the "system integrated response pattern" really the "starting point" of top-down models? Isn't the starting point the delineation of a catchment based on the surface topography assuming a closed water balance? Since most top-down models are calibrated on the streamflow, do you mean streamflow by the term "system integrated response pattern"? Consider clarifying what you mean with the terms, maybe some examples beyond stream flow, and what you mean with "starting point" here.

Reply:

By system integrated response pattern we mean time series of stream flow and other hydrological signatures that can be constructed from these time series (e.g. flow duration curve), or similar variables that characterize the overall flow domain, e.g. solute concentrations in the stream. These system integrated observations are in contrast to point observations of system states, such as *groundwater levels or in-situ soil moisture observations (we do on purpose not mention remote sensing products that claim to provide soil moisture estimates, as it is not clear what these different products actually indirectly estimate and how this information can best be used in models).*

All models need to be based on a definition of the flow domain, i.e. estimates of contributing area or catchment area, and on conservation of mass. This is, however, where the two modelling approaches start to diverge. In this sense, the starting point of top-down approaches is the system integrated data, such as stream flow, when defining the problem as: "We have observations of precipitation input signals and observations of how these input signals are dispersed as stream flow – what is the associated low-pass filter (i.e. model formulation)?". In contrast, the theoretical starting point for bottom-up models is the detailed knowledge of the flow domain and its processes, from which the system integrated response pattern (should) emerge.

We will clarify this in the revised manuscript.

<u>Comment:</u>

Page 5 Line 12: Could you please explain in more detail what you mean with a parsimonious representation of the energy balance?

<u>Reply:</u>

Please see reply to associated comment further above.

<u>Comment:</u>

Section 3 Modelling myths (C1) "Top-down models have a poor physical and theoretical basis": Comparison with Gay-Lussac's law: I think that the comparison with Gay-Lussac's law and the top-down modelling approach is a little misleading. I am not saying top-down models are not physically based. Like most hydrologists I believe that this entire discussion is based on an ill-posed definition and classification of hydrological models into the dichotomy of physically-based and conceptual models. However, with Gay-Lussac's law you can describe the macroscopic state of a system. But you can't say anything about the microscopic state of the system, for example where the molecules really are. Following your arguments and speaking of topdown models now this would mean that you can't say anything about the microscale of a catchment, for example where the water is in your catchment. However, later you argue that you can identify the root zone storage with a top-down model. Is this not part of the microscale? With a macroscopic model you can only infer about the microscale if you constrain the possibilities of the microscale using either additional measurements or process-based reasoning with the help of statistics. However, this is really difficult in hydrology due to the large number of degrees of freedom. For example, if your model is calibrated to mimic the runoff generation and if we assume for a second that the two water worlds proposed by McDonnell (2014) are real, there is no information about the root zone storage in the rainfall-runoff data and it is really difficult to know if what you learn from your models is true.

Overall, it is not clear where you want to go here. A top own model is based on 1.) the conservation of mass and 2.) on the delineation of the landscape into some kind of control volumes mostly in form of a catchment. With a top down model you can hence make assumptions about the macrostate of a catchment or of a similar control volume. With the help of statistics, process-based understanding or additional measurements you might be able to get a grasp of the microscale. So why are you comparing it with a natural law which is constrained by the energy and mass conservation when the model you defend is not? I believe most hydrologists know how a conceptual model works so is this whole comparison necessary at all? Maybe a rigorous definition of macroscale and microscale might help to improve and clarify differences, similarities and linkages between top-down and bottom-up models?

<u>*Reply:*</u>

*We agree with the reviewer that the two systems (gas volume vs. catchment) do have structurally different characteristics. For this reason we understood the comparison as "analogy" (according to the Oxford Dictionary: "partial similarity") and not as full "similarity". We further agree with the reviewer that the microstates are unknown in the gas volume. Similarly, we believe that in principle microstates are, to a certain degree (dependent on the available data and the chosen model resolution/complexity) unknown in a hydrological model (model macrostates rather "integrate […] natural heterogeneity within the model domain […]", p.13,l.33 and elsewhere in the manuscript). A fully lumped, one-process model (e.g. one bucket with a non-linear storage-discharge relationship) would come very close to the conceptualization of a gas-volume. A model that accounts for more individual processes and higher spatial resolution will move away from that situation. Thus, we think that given the roots of conceptual models (e.g. unit hydrograph) the analogy is not too far-fetched. However, we all know that such simple models do not do a good job in representing real world heterogeneity in hydrological systems. The inherent difference between the gas volume (or one-box systems) and catchments (or more detailed models thereof) is that a purely statistical (or data-driven) approach is, following the argument of Dooge (1986), only applicable for systems in the realm of unorganized complexity (i.e. high degree of randomness and complexity). While catchments are too random and complex for an exclusively mechanistic treatment, they are equally not random and complex enough for an exclusively statistical treatment – they rather fall into the realm of organized complexity. In other words, some structure, i.e. distinction of individual processes and/or spatial discretization, is required to meaningfully represent the system. However, within this structure (i.e. within the individual model components, such as the root zone, or, if spatially discretizing, within a given e.g. landuse class) the same principle applies: relatively stable relationships that integrate the sub-domain heterogeneity emerge at the scale of the model domain. In this respect (and for the sake of the argument, assuming no spatial discretization), it is true that the microstates of the root zone cannot be identified and it is <u>not known *where* in a catchment</u> how much water is stored in the root zone. Rather, what is <u>known is that water is stored in the root-zone component</u> and not in e.g. the groundwater component of the system. Therefore, we do not think that the points mentioned by the reviewer point towards a contradiction. Of course the argument can be extended in the same way, when adding spatial discretization, e.g. into land use classes. Each land use class will then be represented by emergent relationships that integrate the sub-domain heterogeneity of this very class – again reflecting the basic idea of knowledge of macrostates without the knowledge of microstates. We will, due to the reorganization of the manuscript, remove the gas laws analogy and provide a clarification for the importance of organization in the revised manuscript.*

Comment:

Page 6 Line 3: The molecular dynamics approach might be untestable and unfeasible but certainly not unnecessary. It is the theoretical basis of the movement of gas particles and hence necessary if you want to understand a system.

Reply:

Agreed, we will reformulate that statement.

Comment:

Page 6 Line 23: Can you please explain in more detail what you mean with parsimonious representation of the energy balance, again?

Reply:

Please see reply to associated comment further above.

Comment:

Page 8 Line 1: Holistic empiricism and on Page 7 Line 6 assign physical meaning to them a priori? Please explain why the two statements are not in contradiction.

Reply:

We think there is some value if the system was seen from the perspective of holistic empiricism – not necessarily that the complete system has to be treated as fully holistic. Assigning parameters derived from observations at the modelling scale and thus assigning physical meaning to individual model components does not contradict the holistic perspective: the parameters obtained from observations at that scale fully integrate the system-internal heterogeneity and its internal interactions. This therefore directly links to holism, which poses that in an interconnected system only sets of hypotheses (i.e. processes at the scale of observations) but not individual hypotheses (i.e. processes at the sub-grid scale) can be meaningfully tested.

The point that distinguishes hydrological systems here is that they are mostly in the realm of organized (i.e. structured) complexity (Dooge, 1986). In other words, they are systems that are characterized by clearly distinct "groups" of processes or components. While it is difficult to reconcile all these components in a truly holistic hypotheses, we believe that each individual component may well be described using the holistic perspective. Thus, we agree, in principle with the reviewer that these two statements are contradicting each other. However, we would argue that this is not the case if the organized nature of catchments is brought into consideration.

Comment:

(C2) "Top down models are too simplistic…" and (C3) "Top-down models are ad-hoc formulations…": Both sections are written clearly and well but I think this has all been said and written down several times. You might consider to shorten this section.

Reply:

This may be true. Yet, the same discussion is coming up over and over again: "Top-down models are too simplistic…" or "Top-down models are ad-hoc formulations…". Therefore we think there is some value in bringing together the loose ends here by analysing the question from both perspectives (which we will try to improve in the revised manuscript)

Comment:

Subsubsection 3.2.2 and 3.2.1: What do you mean with process and spatial complexity. Could you please define complexity and how it relates to the respective models?

Reply:

We use the term "complexity" here to refer to "resolution". Process complexity thus describes, how many individual, interacting processes are considered to generate the response (see example S1). Spatial complexity describes the spatial resolution of the model domain. We will clarify that in the revised manuscript.

Comment:

Page 10 Line 30-31: Is it really multivariate observed response dynamic? At least in one of the cited examples the authors only use streamflow and derivations of it.

Reply:

Good point. What was meant is "multi-objective", which may include both, multiple variables and multiple objective functions. We will reformulate that to be more precise.

Comment:

Page 12 Line 19 - 20: "Competing approaches" Despite the title of the manuscript I had the feeling that the main focus was on defending top-down models. Why do you stress the dichotomy although I understood the overall aim of your opinion paper to be exactly the opposite?

Reply:

Agreed. As mentioned in the reply to one of the comments further above, we now realize that the paper comes across as a simple defence of top-down models. That was not our intention. We tried to resolve the perceived dichotomy between the two model approaches ("all models are physical, all models are conceptual") and we will put some effort to do so in a clearer and more obvious way.

Comment:

Page 13/14 Line 34 / 1-2: I think this sentence is a little misleading. Obviously you can use a DarcyRichards based model on the macroscale. However, you need to use a rather fine discretization of the model elements.

Reply:

Fully agreed, we will reformulate this statement to be more precise.

Comment:

Page 14 Line 16 - 17: Why are you so pessimistic here? Maybe you could add some references so the reader can better understand your pessimism.

Reply:

This is a pointed statement, which was qualified by the stating "in an exaggerated way" (p.14,l.16). However, we think that there is some truth to it, without being pessimistic. We will substantiate the statement with references to work on data/parameter uncertainty (e.g. Beven, Westerberg, McMillan) and uncertainties arising from the model building process (e.g. Gupta, Clark, Wagener).