"*HESS Opinions: The complementary merits of top-down and bottom-up modelling philosophies in hydrology*" by Markus Hrachowitz and Martyn Clark

Reviewed by Hoshin V Gupta, Professor, The University of Arizona

Feb 1, 2017

Summary: Hydrological models exist on a scale of process complexity, from "bottom-up" (BU; often called physically-based) to "top-down" (TD; often called conceptual). The goal of this opinion paper is to improve discussion/exchange between practitioners of these approaches (which are often perceived to be in competition), by a) identifying, discussing and clarifying some common misunderstandings and misinterpretations, b) examining the failures and successes of each approach, and c) highlighting the complementary nature and value of micro-scale process understanding versus the quest for general laws at the catchment scale. The authors discuss some common beliefs such as that "*TD models have a poor physical and theoretical basis*", "*TD models are too simplistic and cannot adequately represent natural heterogeneity*", and "*Top-down models are ad-hoc formulations of untestable hypotheses and always need calibration*".

<u>Based on this discussion one may conclude (here, I am adding my own perspective to the authors statements):</u>

a) There needs to be better recognition of the fact that all hydrological models are to varying extents a blend of "*conceptual*" and "*physically-based*" with each approaching the modeling problem from a different perspective – macroscale versus microscale. With this in mind the authors make the reasonable statement that "*an ideal model would provide an equally good representation of both aspects*".

b) There are at least two major functional behaviors that any model must reflect, which are in regards to how water fluxes are partitioned at the near surface and in the unsaturated root zone. This helps to establish the minimal degree of process complexity that must be represented.

c) Because TD models can, in principle, be implemented with any desired process and spatial detail (depending on availability of data), the TD and BU approaches should really be considered to be complementary, and dependent on the availability of sufficient observations at the selected modeling scale and resolution.

d) Neither approach (TD and BU) can escape the need for rigorous, calibration, testing and post-calibration evaluation, because (i) each approach leads to a hypothesis about the nature of the underlying system, and (ii) direct observations of effective parameters at the relevant modeling scale and resolution are typically not available. In this regard, one must be cautious of using and parameterizing models in such a way as to provide deceptive impression of accuracy (particularly true for the BU approach).

e) It would be desirable to establish a mainstream culture of rigorous testing of alternative model formulations (hypotheses), robust model calibration and, and systematic assessment of model uncertainties, based on the establishment of some "*minimum requirements*".

f) Progress in catchment-scale understanding of hydrological functioning and the related development of models for more reliable predictions will be well served by (i) a better understanding of how natural heterogeneities at all scales aggregate to larger scales and how this influences the hydrological response, and (ii) a convergence of TD and BU strategies, in particular with respect to exploiting the features of organization in these complex systems in a hierarchical way. A potential way forward is via large sample, comparative hydrology to identify patterns and generally applicable, functional relationships.

My Review Comments: I found little in the substance of this opinion paper to disagree with. My main comments have, therefore, to do with the fact that the presentation tends (I suspect partly unintentionally) to come across as a defense of the TD approach, rather than a balanced evaluation of the strengths and weakness, and complementary nature, of the TD and BU approaches. Certainly in the Gupta et al (WRR 2012, Model Structural Adequacy) paper, of which Clark is a co-author, we argued for the commonality of underlying structure of most if not all hydrological models based on the steps involved in model building, and the need for more cross-fertilization across the modeling community. I very much like the fact that the authors of this paper emphasize the issues of the perceived (but unnecessary) conflict between the TD and BU approaches, but I feel that the argument could be refined and made more balanced by taking note of the fact that many of the points raised in defense of TD modeling are really more general comments that apply to all levels of model complexity – from BU to TD, and revising many of the concluding comments appropriately.

Below, I provide the summary I prepared (of major points presented) while reviewing the paper. While doing so, I found myself generalizing some of the comments made to extend to both TD and BU modeling, and slightly reorganizing the concluding comments. I provide them here in case it helps the authors to see these remarks from a slightly different perspective, and thereby to be useful in strengthening the paper.

In conclusion, I commend the authors on a very nice commentary ☺.

Summary of the Paper:

The issue:

- Hydrological models frequently fail to reproduce the hydrological response in periods they have not been calibrated for, thereby providing unreliable predictions – this suggests that some of the underlying processes that control how water and energy are stored in, transferred through, and released are not sufficiently well represented.
- Such models exist on a scale of process complexity, from "bottom-up" (BU; or physically-based) to "top-down" (TD; or conceptual).
    - Bottom-up refers to detailed, high-resolution descriptions of small-scale processes numerically integrated to larger (spatial) scales, that include detailed and explicit treatment of conservation of mass, energy and momentum, and whose parameterizations are based directly on observations of fluxes on the small scale.
    - Top-down refers to less detailed representations, often spatially lumped at catchment-scale, whose treatment of conservation and parameterizations are less detailed.

- Unfortunately, there is little fruitful exchange between BU and TD modeling communities, and communication is often limited to mutually highlighting each other's deficiencies. To achieve progress it is important that we:
  - Examine the failures and successes of each approach
  - Appreciate the complementary nature and value of micro-scale process understanding versus the quest for general laws at the catchment scale
- Goal of this paper:
  - To identify, discuss and clarify common misunderstandings and misinterpretations of competing modeling approaches
  - To provide a perspective of how to take advantage of different modeling philosophies so as to improve predictions

Modelling philosophies:

- The BU strategy has two main features; (1) Explicitly accounts for spatial heterogeneity, and (2) Provides a rigorous and physically consistent way to encapsulate and formalize theoretical knowledge of dominant processes. In principle, this:
  - Enables hypotheses that can be individually scrutinized and tested against observations
  - Provides meaningful representation of natural feedbacks between individual parts of the system -- implementations of this inductive approach to hydrology have the potential to reproduce emergent patterns.
- The TD strategy is based on representing emergent, *system-integrated* (catchment scale), response patterns reflected by information in available data (mainly limited to areal estimates of precipitation, potential evaporation and stream flow). It aims to reproduce the observed dispersion patterns that reflect catchment-internal organization and feedback. While maintaining mass balance and parsimonious representation of the energy balance, it does not explicitly define and describe the detail of underlying processes.

Discussion of Typical Modelling Beliefs:

- **Top-down models have a poor physical and theoretical basis**
  - While having an element of truth, it is demonstrably clear that observation-based, functional relationships at the macroscale can be valuable descriptors at the macroscale without loss of essential information.
  - Since water flows in a catchment follow the observable, physical phenomenon of dispersion, controlled by water and energy input, gravity, and flow resistances, TD model development is the attempt to identify functional input-output relationships between that emerging through organization at the catchment scale.
  - Accordingly, it should be possible to test competing hypotheses without resorting to small-scale physics. By examining a large enough sample of systems, the emerging patterns and associated (tested) functional relationships can facilitate similarity analysis and classification, and facilitate the "*search for general laws at the macroscale*".
  - From this perspective, TD models can be considered as "physically-based" and parsimonious representations at the macroscale. Given that they consist of large-scale conservation equations, they necessarily require large-scale flux

parameterizations (i.e. the closure problem) that act on system-average water quantities. The challenge is to assign clear physical meaning.

- We argue, therefore, that well implemented and tested TD models can have a robust physical and theoretical basis, and that it is possible to relate the structure of top-down models to stores and fluxes in nature, albeit at a different spatial scale and process resolution than bottom-up models.
- There is, however, room for improvement, such as (a) providing an explicit physically consistent treatment of energy and momentum balances, and (b) determining the level of detail at which dominant catchment processes must be resolved to reproduce the observed system response in a meaningful way.

- **Top-down models are too simplistic and cannot adequately represent natural heterogeneity**
  - It is true that TD models are often assumed, a priori, to be meaningful representations of the system, rather than treated as hypotheses to be tested. While displaying high skill with regard to a calibration objective, they can struggle to reproduce additional system internal dynamics and emerging patterns, resulting in poor predictive power.
  - While lack of adequate model calibration, testing and evaluation partly arises both from (a) insufficient exploitation of information in available data, and (b) lack of suitable data to more effectively constrain models, many models remain ill-posed inverse problems.
  - Two issues are of importance here – Process and Spatial complexity:
    - *Process Complexity:* All models of terrestrial hydrological systems must represent the two major types of partitioning that control how water is stored in and released through upward, downward or lateral fluxes. The first is the near-surface partitioning of precipitation into (i) evaporation and sublimation, (ii) overland flow and (iii) infiltration. The second is the partitioning of infiltrated water into (iv) soil evaporation, (v) plant transpiration, (vi) shallow, lateral subsurface flow, and (vii) percolation. Regardless of (TD or BU) strategy, the level of detail for resolving individual processes at the two partitioning points must be decided. Questions to be answered are: *How much detail is necessary to reproduce observed dynamics and pattern? How much detail is warranted by the available data to meaningfully parameterize and test the chosen process representation?* As long as simplification encapsulates relevant dynamics of the system, lumping does not necessarily involve loss of information, and has the potential to provide a description that is consistent with real world observations at the scale of interest.
    - *Spatial complexity:* The degree of spatial complexity that can be incorporated in a model depends on the detail of available information (which types of heterogeneity are present, how do they affect water storage and release, and which can be captured by a single emergent functional relationship?). In principle, TD models can be formulated at any level of process and spatial complexity, limited only by available information. So the decision regarding which level of detailed process representation is feasible/necessary must be made by the modeler on basis of the available observations. It is also important to ask whether BU models may actually be deceptive about the accuracy that is

implied by their formulation, when appropriate observations at the scale and resolution of interest are not available.

- **Top-down models are ad-hoc formulations of untestable hypotheses and always need calibration**
  - It is true that applications of TD models often assume their appropriateness rather than being treated as hypotheses to be tested. More correctly, different environmental conditions should dictate the specification of prior parameter information and/or process inclusion/activation.
  - Further, *all* models (TD or BU) are aggregations of distinct hypotheses that should ideally be tested independently to avoid the adverse effects of equifinality. The question is "*To which level do model components then have to be disaggregated to be constitute testable hypotheses*?" Inevitably, any treatment of any model as a single hypothesis is likely to remain a weak test. In contrast, individual testing of system sub-components provides more information because the sub-components are necessarily less complex than the overall model and results in a stricter test.
  - So this is not a limitation specific to TU models, and the main challenge is availability of relevant observations at the scale and/or resolution of interest, which gives rise to the need for calibration. Meanwhile, weak model tests are preferable to no tests at all.

Implications, potential ways forward and concluding remarks:

- All hydrological models are to some extent "conceptual" and to some extent "physical"; they largely only differ in the degree of detail they resolve the system, which in turn is dictated by the available data. TD models approach the problem from a macroscale understanding, while BU models emphasize the micro-scale perspective. Needless to say, an ideal model would provide an equally good representation of both aspects.
- TD and BU modeling are complementary, with TD modeling constituting a physical-based approach to hydrological modeling that can be firmly rooted in holistic empiricism, similar to statistical physics.
- TD models can, in principle, be implemented with any desired process and spatial detail (depending on availability of data).
- TD and BU models must both reflect our conceptual understanding of the system in regards to how water fluxes are partitioned at the near surface and in the unsaturated root zone. Since all the relevant fluxes can be present in any environment (with different relative importance), all TD top-down models must have the same fundamental model structure (but not necessarily the same parameterization) to reflect these processes.
- In the absence of sufficient observations at the modeling scale and resolution, all hydrological models (TD and BU) remain hypotheses and require rigorous, testing and post-calibration evaluation. The fundamental problems in catchment modeling do not lie in the type of model used, but rather in the way a model is applied.
- Progress in catchment-scale understanding of hydrological functioning and the related development of models for more reliable predictions hinges on a better understanding of how natural heterogeneities at all scales aggregate to larger scales and how this influences the hydrological response. Although acknowledged in the early 1980s,

remarkably little progress was made. A potential way forward is large sample, comparative hydrology to identify pattern and generally applicable, functional relationships.

- All hydrological models (TD and BU) applied at scales beyond the plot scale require some degree of calibration, as direct observations of effective parameters at these modeling scales and resolutions are typically not available. Improper application of BU model parameters from observations that do not match the modeling scale and/or resolution can lead to deceptive impression of accuracy.
- If both TD and BU models are to be useful as scientific tools, a mainstream culture of robust model calibration (and post-calibration), rigorous testing of alternative model formulations (i.e. hypotheses), and systematic assessment of model uncertainties (including parameterization) must be achieved. This should involve evaluation with respect to multiple variables and model states, including (data permitting) model sub-components, as well as to multiple criteria, and the establishment of "minimum requirements".
- Taken together, these arguments suggest that somewhat arbitrary dichotomy between TD and BU models tends to lead to some degree of confusion. In reality, all models are to some degree conceptual. The real issues have to do with:
  - The inadequate ways in which models are applied. Recognizing this may help shift the focus towards the real fundamental questions in catchment-scale modeling, including "how much detail do we need in our models?" and "how much detail is warranted by data?"
  - Finding a balance that allows us to best describe the system based on scientifically robust grounds
  - Accepting a more rigorous culture of model testing, to adjust process and spatial complexity to the environmental conditions and data availability in specific catchments, so as to reduce the risk for oversimplifications and system misrepresentations (TD) while embracing the value of zooming out and making use of emergent processes (BU).
- On balance, we believe that improved hydrological understanding will require a convergence of TD and BU strategies, in particular with respect to exploiting the features of organization in these complex systems in a hierarchical way. We strongly encourage researchers to both acknowledge and actively use the advantages of both modeling strategies in order to strengthen the science of hydrology.