Response to Anonymous Referee 1

REPLY TO GENERAL COMMENTS:

"The manuscript describes a modified curve number (CN) used to delineate hydrological response units for hydrological modeling.

The topic is of interest for the hydrological community, because although a huge amount of literature exists on HRUs, how to best subdivide the catchment for hydrological modeling remains an open question."

We thank the Anonymous Reviewer (herein referred to as AR1) for his detailed review, with constructive comments and suggestions. We appreciate for finding this work useful for the hydrological community and within the state-of-the art interests.

I am not so convinced that the results of the study show the practical utility of the method. Therefore, I recommend expanding the case study (see detailed comments).

Thank you for your constructive critique. Based on your recommendations, as well as the suggestions by other reviewers, we have expanded our analyses, to make the case study more convincing. Please, refer to our response to your specific comment No 6, where we present the outline of the expanded analyses.

I also would not refer to the modified CN as a "new framework". The author propose a different definition of the CN, which as a result, becomes something different from the CN, and therefore can also be given a different name.

In the revised article, we will remove the characterization of our technique as "framework", as suggested. On the other hand, we disagree with your statement about the *different definition of the CN*. The key aim of our paper was to discuss suitable practices for delineating HRUs within hydrological modelling, on the basis of distributed CN information. To facilitate the production of CN maps (which is essential for the delineation of HRUs), we have also proposed a GIS-based procedure that uses the empirical formula (1) to estimate CNs at the grid cell scale, according to permeability, land cover and slope classes. In this procedure, the novelty refers to the empirical formula itself, not to the resulting quantity, which is the widely known curve number parameter.

In fact, this formula was derived within a recent research project, involving the collection of flood data across a number of experimental basins in Greece and Cyprus (Efstratiadis *et al.*, 2014). At each basin, some dozens of flood events were analyzed, to provide, among other things, representative CN values. The proposed procedure ensured good fitting to the empirically estimated CNs at the basin (or sub-basin) scale. An interesting conclusion was the little lower values of optimized CNs relative to the values obtained by the classical SCS procedure that is based on soil and land use data. For this reason, we decided to increase the number of soil classes from four to five, on the one hand, and to employ the slope as additional explanatory variable. We note that the weight assigned to the slope is 1/3 of the weight assigned to permeability.

In the revised version, we will add a paragraph providing synoptic information on the derivation of the proposed formula and its possible range of applicability.

As the authors are considering the suitability of the CN in the optic of a specific application (i.e. the subdivision of catchment in HRUs for a specific hydrological model), it is unfair to speak *about limitations of the CN, and of "improvements". The CN was not developed with the authors' application in mind."*

Obviously, the initial objective of SCS when introducing the CN concept, many decades ago, was totally different, i.e. to predict surface runoff from gross rainfall. The CN parameter has some very attractive characteristics, since it allows quantifying the flood response of a river basin by means of a dimensionless quantity. It is well-known that CN is associated with the characteristic parameter of the SCS method the potential maximum retention, S (in mm), through the empirical relationship $S = 254$ (CN/100 – 1).

In the context of our work, the CN concept is expanded to be considered as the "hydrological identity" of an area applicable at any spatial scale (from the grid cell to the basin scale). Of course, in contrast to the SCS method, here, the value of CN (more specifically, the average value across the area of interest) does not correspond directly to any of the model parameters. However, the relative values of CNs can be used to identify suitable and physically-consistent parameter ranges within the calibration procedure.

In the reviewed version of the paper, we have already provided some hints on the above issues. For instance, in the conclusions we write (p. 27, lines 20-24):

"A raster map of CN values may have several applications, since this parameter is commonly used in watershed hydrology. For instance, it can provide information on potential maximum soil moisture retention at the grid-scale, since CN is directly associated with this key property of soils. Another application, emphasized in our research, is the use of CN data as proxy for delineating HRUs in the context of distributed hydrological models."

Further clarification will be provided in the revised document, to better explain the use of the curve number concept within our methodology.

REPLY TO SPECIFIC COMMENTS:

1) Abstract. I did not see evidence in the results that support the claims made in the abstract.

1a) The authors state that the new approach aims at reducing the subjectivity introduced by the definition of HRUs. Yet, the paper does not provide clear support on how to select the number of HRUs. The authors propose (line 7, page 11) that "the number of HRUs equals the number of discharge monitoring stations". This is however a general recommendation given without support, and cannot be considered a general conclusion of the study.

As correctly mentioned by AR1, in p. 11, lines 6-10 we provide the following recommendation:

"As a general principle in the HRU configuration procedure, it is proposed that the number of HRUs equals the number of discharge monitoring stations (provided, obviously, that both the quantity and quality of hydrometric data are satisfactory). Therefore, if N discharge monitoring stations are available, then up to N CN parameter classes should be considered, in order to formulate the corresponding HRU classes."

This recommendation is not a new "idea". A number of significant research articles dealing with best practices in hydrological modelling accept that only a small number of parameters can be identified on the basis of a single observed response of the hydrological system. One of those pioneering works is the article by Jakeman and Hornberger (1993), who investigated numerous parameterizations, concluding that (p. 2646):

"The major result of our study is that, for catchments in temperate climates but over a tremendously wide range of scales, only a handful of parameters can be reliably estimated from rainfall-runoff data".

In their classical article, Wagener *et al.* (2001), discussing state-of-the-art approaches in the development and application of hydrological models, state (p. 14):

"The level of structural complexity actually supported by the information contained within the observations is defined here simply as the number of parameters that can be identified. … Results from previous research suggest that, in the case of rainfall-runoff modelling, up to five or six parameters only can be identified from time-series of external system variables (i.e. streamflow and rainfall) using traditional single-objective calibration schemes (e.g. Wheater et al., 1986; Beven, 1989; Jakeman and Hornberger, 1993; Ye et al., 1997)."

In an attempt to improve the identifiability of parameters in case of complex schemes (e.g., semi-distributed), Efstratiadis and Koutsoyiannis (2010) have proposed the introduction of multiple criteria, each one explaining a "handful of parameters", i.e.:

"… extending the empirical rule expressed for lumped models, we should retain a ratio of about 1:5 to 1:6 between the number of criteria and the number of parameters to optimize, to provide a parsimonious representation of the multi-objective calibration problem."

However, not all researchers confirm this rule, which is reasonable, since this rule does not have any strict theoretical justification. For instance, in a recent article, Fenicia *et al.* (2016; thank you for suggesting this very important article) found that much fewer parameters were essential to represent ten observed hydrographs across a river basin. Discussing the results of their analysis, they concluded that:

"It is well known that, when calibrating hydrological models on individual hydrographs, only a ''handful'' of parameters are often identifiable [e.g., Jakeman and Hornberger, 1993]. Less clear is what would be an appropriate level of model complexity for a distributed hydrological model. In our previous work, we found that five to eight parameters are sufficient to reliably capture the hydrographs of some of the case study catchments [Fenicia et al., 2014]. If these numbers scaled up linearly, a tenfold amount of model parameters would be necessary to simulate 10 hydrographs using a distributed model. In this study, where we attempted to simulate 10 seemingly different hydrographs, two ''handfuls'' (11 parameters) were found to be sufficient. This result may give a broad indication of the complexity of a distributed model in similar conditions of application, yet care must be taken because this result is derived solely based on case study location data."

It is important to mention that Fenicia *et al.* (2016) recognize that their outcomes cannot be generalized, since they are derived from a specific study. Nevertheless, their outcomes are very interesting and will be taken into consideration in the revised article.

Regarding our recommendation of the one-to-one correspondence of HRUs and observed model responses, we will better highlight that this is an empirical hypothesis, but one that is quite well-supported by the common hydrological experience. In addition, within the case study, we will enhance our analyses, as proposed by all reviewers, to better justify the selection of three HRUs in order to calibrate the model against three observed hydrographs. For details, please, refer to our response to your comment #6, regarding the expansion of the case study.

1b) The authors state that "the CN-based parameterization allows the user to assign as many parameters as can be supported by the available hydrological information". I don't see how this would be a distinctive outcome of the study. One can always get more HRUs by overlaying different properties such as geology, topography, soils, etc.

Key aim in our methodology is avoiding subjectivity on the formulation of HRUs, based on any type of geographical data. In this respect, we favor the use of the well-known CN concept, which allows providing HRUs through appropriate classification of the raw (i.e. raster-type) CN map.

Therefore, the number of HRUs can be any integer within the potential range from a single HRU (lumped case) to Nmax HRUs (fully-distributed case), where Nmax is the product of all available combinations of soil permeability, land cover and terrain slope classes (at most 5 × 5 × 5 = 225 classes). On the other hand, by overlaying, e.g., N1 classes of characteristic 1 and N2 classes of characteristic 2, and so on, one will by definition obtain the product of all available classes (in the specific case N1×N2 HRUs). We note that this product is formally known as common refinement of the partitions, while in the GIS terminology the related procedure is often called "union of layers" (Efstratiadis *et al.*, 2008). In our opinion, the flexibility of delineating as many HRUs as desirable, irrespective of the background geographical information, is a strong advantage of the proposed parameterization procedure, and will be emphasized in the revised article.

1c) "the CN-based parameterization reduces the effort for model calibration". Surely, this depends on how many HRUs are selected, and the number of HRUs can vary.

For exactly this reason we recommend using as many HRUs as the number of available hydrological information, thus ensuring an optimal equilibrium between the representation of process heterogeneity and model complexity, also reflected in terms of time effort in calibration.

2) Introduction. I found the introduction quite imprecise in the terminology, and incomplete concerning the coverage of the relevant literature e.g. [Kampf and Burges, 2007], which talk about different discretization approaches, [Fenicia et al., 2016], which deal with the problem of how to discretize the catchment into HRUs, a discussion of alternative, more common approaches to define HRUs [e.g. Scherrer and Naef, 2003].

Thank you for this remark and the recommended articles. In the revised document, we will expand the literature review, including the proposed citations.

2a) For example, when talking about the differences between lumped to distributed models, several imprecise statements are made. E.g. that lumped models are simple (compared to what? To all distributed models? Not necessarily true), model parameters are associated with the macroscopic properties of the watershed (not necessarily true), lumped models have limited physical background (not necessarily true), model parameters need to be inferred by calibration (not necessarily true). With respect to distributed models, the authors state that the model domain is discretized in "finely resolved" computational units (not necessarily true), that in theory parameters are estimated from field data (not necessarily true), etc.

Our comments concerning the differences between lumped, semi-distributed and fully-distributed hydrological models, as well as the watershed subdivision approaches, summarize some generally accepted assumptions of the hydrological community (initiating from Beven's 1989 breakthrough paper). We recognize that not all of them are 100% valid under all possible conditions, but we can agree that they are consistent with the broader hydrological experience. In particular:

Statement 1, "lumped models are simple": The vast majority of lumped models are simple, in the sense that they aim to describe the key processes at the river basin scale using few conceptual equations and a small number of parameters. Of course, one can provide lumped models of much increased complexity, but definitely this is not the usual case.

Statement 2, "[lumped] model parameters are associated with the macroscopic properties of the watershed": Since lumped conceptual models are by definition applicable to the macroscale, their parameters represent "average" properties of the basin (e.g., average soil capacity, average hydraulic conductivity). Of course, the values of average parameters cannot be associated with real field quantities, while some parameters may only have empirical evidence.

Statement 3, "lumped models have limited physical background": To our opinion, this is expected, since model parameters refer to the macroscale, while the theoretical equations of hydraulics refer to the point scale.

Statement 4, "[lumped] model parameters need to be inferred by calibration": As already mentioned, we emphasize the macroscale. At the macroscale, there is no direct correspondence of lumped model parameters with small-scale physical properties; "parameters" have to be inferred, while "properties" can be identified (under several premises) through field measurements. An alternative to calibration are regionalization approaches, which are in fact empirical relationships that have been also calibrated against observed data.

Statement 5, "[in distributed models] the model domain is discretized in "finely resolved" computational units": We agree that, quite often, the computational units of distributed models may be relatively large, e.g. up to some km^2. However, the spatial scale of such units with may still differ by one or more orders of magnitude compared to the basin scale, thus justifying the term "finely resolved", at least from the comparative viewpoint.

Statement 6, "parameters [of distributed models] are estimated from field data": Actually, several times (maybe, almost always) the parameters of distributed models are treated as free variables, to be inferred through calibration. Thus, this statement will be rewritten to better reflect the reality.

2b) Considering the focus on HRUs, I would have expected a broader coverage of the relevant literature, e.g. [Kampf and Burges, 2007], which talk about different discretization approaches, [Fenicia et al., 2016], which deal with the problem of how to discretize the catchment into HRUs, a discussion of alternative, more common approaches to define HRUs [e.g. Scherrer and Naef, 2003].

Thank you for this remark. We will study further the literature on HRU delineation as well as the publications that you recommend, and incorporate them in the introduction in the revised paper.

3) Scope of research. The first paragraph gives a definition of the CN. It is redundant with what described in paragraph 4, and creates confusion, as before defining the modified CN approach one would need to recall the original CN. Anyway, this paragraph does not belong in a scope section.

Thank you for this remark. According to suggestions by all reviewers, we will substantially reorganize the material provided in the first two sections, including removing redundant material, in order to better highlight the key methodological issues of our research and the major questions to be answered, and therefore improve the readability of the paper.

Lines 6-11 page 5 are not scope, so they don't belong here. They could have been conclusions, but they are not, because they are not supported by the analysis.

Thank you for your recommendation. Actually, conclusions should be moved to the closing section.

4) The motivation for the modified CN is quite weak and buried in the text.

Thank you for this remark. In the revised text, we will better explain our motivation, key issues of which are: (a) improving the estimation of CN, by considering five instead of four soil types, as well as the drainage capacity, within its definition; (b) providing an empirical formula that allows for crating raster maps of CN on the basis of distributed classes of permeability, land cover and terrain slope; (c) taking advantage of the classified CN map to delineate HRUs, in the context of distributed hydrological models.

4a) Line 21 in page 7 states that an important shortcoming of the standard CN method is that "it does not take into account the effect of slope". The authors have obviously in mind a specific application, for which slope would be important. It would be useful to reveal this application to the reader in a clear way, before talking about limitations. And as stated in my general comments, it would be useful to state whether such application motivated the development of the CN, otherwise, it is unfair to talk about limitations of the CN tout court.

The original CN values within SCS experiments were identified on the basis of observed rainfall-runoff events in small agricultural watersheds with generally mild slopes. For this reason, some researchers have proposed empirical formulas for adjusting slope to CN, thus resulting in much improved runoff simulations (e.g., Huang *et al.*, 2006).

4b) Line 24 in the same page states that "steep slopes cause reduction of initial abstractions, decrease of infiltration, and reduction of the recession time of overland flow". I am not sure this is true in general. For example, on sandy hillslopes, there is no lateral flow. Such statements need to be supported with some references and need to be given a perspective.

There are numerous research articles investigating the runoff dynamics in steep landscapes. As example, we will cite the work by Montgomery and Dietrich (2002), also containing an extended list of articles dealing with the above topic.

Regarding your comment about lateral flow generation in sandy hillslopes, we note that in our proposed formula (1), the weight assigned to soil permeability is three times the weight assigned to drainage capacity, thus recognizing the increased importance of permeability relative to slope in runoff generation. Nevertheless, due to erosion processes, HRUs resulting from combinations of sandy soils and very steep slopes should be rare.

5) section 4.2 is titled "Novel GIS-based framework for CN estimation". It is unnecessary to write "novel". It goes by itself that if you write a paper, you present something novel, otherwise there is no point in writing a paper. In addition, it is unnecessary to write "framework" for something that is simply a different formula for the CN. Probably, it is also incorrect to write CN, as being a different formulation, it does not need to be called CN anymore.

In fact, the proposed methodology is not just a different formula for CN calculation. Although a new formula is proposed (which is original, since it introduces additional permeability classes types, as well as the introduction of drainage capacity as additional explanatory variable for runoff generation), the modified CN approach is expanded further. For, the CN parameter is calculated for every unique combination of the three physiographic characteristics (soil permeability, vegetation density, drainage capacity), resulting in a distributed basin map of CN values. The map is then used to delineate the desired number of HRUs that will be used as an input in any type of hydrological models, i.e. from lumped (extreme case of a single HRU) to semi-distributed (e.g. the HYDROGEIOS case) as well as fully-distributed (considering the raster map of CNs identical to the HRU map).

According to your suggestion, we will remove term "framework" (maybe the term method or methodology is more appropriate), at the same time better highlighting the novelties of our approach.

6) The case study should be considerably expanded.

As also explained in detail in our response to Reviewer 3, in the revised article we will demonstrate the outcomes of a much extended analysis, involving a calibration experiment with one to five HRUs calibrations. Our objective is to better support our fundamental hypothesis that the number of HRUs should equal the number of available response time series across the basin. Details are provided in the comment after the next one.

6a) Are the discharge stations producing different hydrograph? Can this be shown through some signatures?

In the revised text we will quote typical flow signatures, in terms of mean flows and characteristic percentiles from the flow duration curves, to allow quantifying the differences between the three hydrographs. However, since the flow records at the three monitoring stations only capture a three year period, and the time series do not overlap during this period, these signatures will not be fully representative, thus they should be interpreted carefully.

6b) Is the model comparison meaningful? For example, it would be useful to see how the different HRU discretization compare to a "null hypothesis" where 1 HRU is considered, particularly since results seem to favor the discretization in the smallest number of HRUs.

As mentioned, we are currently performing additional analyses with various parameterizations, from one to five HRUs calibrations. In this vein, the CN parameter map of Nedontas river basin, initially containing 34 classes (Fig. 6, left), is used to delineate 1, 2, 3 (the number of HRUs proposed and used in the initial simulations), 4 and 5 HRUs. We remind that in HYDROGEIOS, 7 parameters are assigned to each HRU, thus, for each configuration the total number of control variables is $7N$, where N is the number of HRUs. Therefore, the simplest parameterization, the lumped (homogenous) basin, comprises 7 parameters, while the scenario with 5 HRUs involves the estimation of 35 parameters. In all cases, the same objective function is optimized, comprising efficiency metrics at the 3 monitoring stations, and other terms that allow testing realistic responses for the remaining processes.

It is important to note that in this computational experiment we did not allow manual interventions within calibration, as done in the context of the scenarios examined in the initially submitted manuscript. We simply run the global optimization algorithm assuming the full set of parameters (7, 14, 21, 28 or 35) within their physical bounds. In order to take into account the increasing computational burden of optimization against the number of control variables, we set a maximum budget/limit of $2000N$ trials, where N is the number of HRUs. Thus, for the lumped approach we allowed up to 2000 function evaluations in each calibration scenario, while for the more complex configuration ($N = 5$) the budget/limit was increased to 10000 function evaluations.

To ensure unbiased results as much as possible, all calibrations were carried out by considering the same initial conditions and by assigning the same parameter values for the other model components (groundwater conductivities, sub-basin routing rates). Moreover, to avoid getting trapped in a local optimum, thus resulting in sub-optimal model performance, we employed several independent optimizations for each parameterization, and kept the best solution.

As will be demonstrated in the revised article, the above experiment further confirms our fundamental hypothesis that the "optimal" balance, in terms of model performance against computational effort, is ensured by considering the parameterization with three HRUs, which equals the number of available hydrographs.

6c) What is the value of the objective function?

The objective function comprises several criteria that are weighted in a single performance measure; these also include penalty terms for prohibiting the generation of unrealistic patterns for simulated groundwater levels, penalty terms for intermittencies etc. (see p. 14, lines 9-12). In Table 4, we only list efficiency and high-flow efficiency values that are easy to understand (hydrologist are very familiar with this statistical metric), while we avoid showing the total value of the objective function, which is an aggregated measure without any physical or mathematical interpretation.

Do more complex models perform better in calibration? It should be so, otherwise there may be a problem in the calibration process, also given the huge number of parameters for the models with large number of HRUs.

In theory, as the number of parameters increases, models become more flexible, thus they can be better fitted to the observed data. However, the broad hydrological experience has shown that this is not true, highlighting that more parsimonious models are more robust against uncertainties. A well-known reason of the relatively poor performance of too complex schemes is associated with calibration, given that even the most advanced optimization methods fail to ensure the "globally" optimal fitting. Even if the increase of model parameters (or, equivalently, the number of HRUs) results to improved calibrations, this usually requires significant sacrifices in terms of time resources. The above issues are already demonstrated in the current case study, which will be further extended in the context of the revised article.

6d) Are the models able to capture the spatial variability in observed responses? E.g. are they able to reproduce the different signatures?

Thank you for this useful comment. As mentioned in our response to comment 6a, in the revised version, we are planning to use some typical flow signatures for the observed hydrographs and compare them against the corresponding signatures of the simulated data, to ensure a more comprehensive assessment of the predictive capacity of the model.

7) In the summary and conclusion section, I had difficulty in identifying the conclusions. It would be better to skip the summary (anyway the abstract is already a summary), and just focus clearly on the conclusions.

Thank you for this remark. In the revised document we will re-organize the conclusions section, so that the answers to the research questions are more clear and better understood by the reader.

References

Beven, K. J.: Changing ideas in hydrology – The case of physically-based models, *J. Hydrol.*, 105, 157-172, 1989.

Efstratiadis, A., Koukouvinos, A., Dimitriadis, P., Rozos, E. and Koussis, A. D.: Theoretical documentation of hydrological-hydraulic simulation model, *DEUCALION - Assessment of flood flows in Greece under conditions of hydroclimatic variability: Development of physically-established conceptual-probabilistic framework and computational tools*, Department of Water Resources and Environmental Engineering - National Technical University of Athens, National Observatory of Athens, Athens, 2014.

Efstratiadis, A., Nalbantis, I., Koukouvinos, A., Rozos, E., and Koutsoyiannis, D.: HYDROGEIOS: a semi-distributed GIS-based hydrological model for modified river basins, *Hydrol. Earth Syst. Sci.*, 12, 989-1006, doi:10.5194/hess-12-989-2008, 2008.

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *Water Resour. Res.*, 52, 954–989, doi:10.1002/2015WR017398, 2016.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: Is there a correspondence?, *Hydrol. Processes*, 28(4), 2451–2467, doi:10.1002/hyp.9726, 2014.

Huang, M., Gallichand, J., Wang, Z., and Goulet, M.: A modification to the soil conservation service curve number method for steep slopes in the Loess plateau of China, *Hydrol. Process.*, 20(3), 579–589, 2006.

Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, 29(8), 2637–2649, doi:10.1029/93WR00877, 1993.

Montgomery, D. R., and Dietrich, W. E.: Runoff generation in a steep, soil-mantled landscape, *Water Resour. Res.*, 38(9), 1168, doi:10.1029/2001WR000822, 2002.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S.: A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5(1), 13–26, doi:10.5194/hess-5-13-2001, 2001.

Wheater, H. S., Bishop, K. H. and Beck, M. B.: The identification of conceptual hydrological models for surface water acidification, *Hydrol. Process.*, 1, 89-109, 1986.

Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M. and Jakeman, A. J.: Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments, *Water Resour. Res.*, 33, 153-166, 1997.