General analysis: First of all, I want to apologize because the Buendia et al. (2016) paper that states “Overall, results have indicated that increased forest areas are the major driver of reduced streamflows and the magnitude of peak floods” is not the Buendia et al. (2016) paper quoted by the authors but another one (http://dx.doi.org/10.1016/j.scitotenv.2015.07.005) published two months later studying another basin. Indeed, the paper quoted by the authors found that for the embracing studied catchment (Talarn), somewhat less than 50% of the runoff reduction (37%) could be attributed to forest cover encroachment, but these authors state that “Neglecting re-vegetation could lead to erroneous projections resulting in an underestimation of the runoff future trends; thus, evolution of forested cover should not be ignored when designing land and river basin management plans at the light of global change scenarios”. Therefore, this paper should not be fairly cited as a reason to omit the role of land cover change in streamflow temporal trend studies.

I want to acknowledge the effort made by the authors to follow my recommendations. The new information provided is noteworthy but not easy to understand, so I am trying to analyse it and to provide updated recommendations to the authors.

Trends: The data shown in the table S1 are really striking as they point to relevant internal inconsistencies in the model results. First, the fact that 70% of the gauging stations show positive or negative significant residual trends, with a coefficient of variation of 822%, demonstrates a high uncertainty of the model results. Second, the spatial distribution of these trends in figure S1 shows disordered patterns; some successive gauging stations without any significant tributary between them show opposite trends, such as Castellbell and Abrera stations on the Llobregat River and the two Les Masies de Roda stations on the Ter River. Third, some of the plots in figure S2 show very asymmetric abnormal shapes, either positive (Balsareny, Fogars de la Selva (Pont Eiffel), Sallent) or negative (Guixers, Sant Feliu de Buixalleu).

Negative trends of the residuals may be attributed to the role of increased forest cover in the area, a very likely behaviour already demonstrated by previous works, but here it is more difficult to attribute positive trends to hydrological reasons. I wonder whether the calibration-validation strategy used (randomly selected for each station independently) may lead to different sensitivities of the model to climate forcing at each station in such a varied climate and induce this scatter. As Buendia et al (2016) stated “precipitation appears to follow a generalised decreasing trend, although the significance of these results depended strongly on the time period considered”. In fact, the authors do not provide any evidence of the validity of their consideration (line 371): “We consider that, given the high number of gauging stations in our study, randomly determining the calibration and validation periods for each station effectively captures all spatio-temporal variability in both periods. Thus, a bootstrapping method to repeatedly resample the calibration and validation periods, which is time consuming and would imply running many more iterations, is not necessary”.

SWAT+: This is not a physically-based model even if it can provide with good results. This qualification of this model contributes to the degradation and loss of usefulness of terminology and concepts. The addition of complementary processes does not modify its essentially empirical character. There is an agreed methodological caveat that just because a model gives good results does not imply that it is for the good reasons (in particular, structure, internal stores and flows). In fact, the results shown by the authors in table S1 may provide a corollary of this principle: despite acceptable flow calibration/validation tests, residual discharge trends show chaotic values difficult to attribute to hydrological reasons.

Recommendations:

In general terms, the manuscript describes the modelling exercise, shows its results, makes some comparisons with observed data and claims the success of the exercise as it "led to successfully simulating hydrological and anthropogenic processes in water-scarce Mediterranean basins" and "resulted in notable improvements in hydrological modelling and its potential use to support decision-making in the water management sector" without contributing no evidence of these successes and improvements.

The authors should not claim good modelling results beyond acceptable tests of efficiency and uncertainty, but should be much more analytical by discussing their strengths and weaknesses and suggesting ways to remedy the latter.

- The authors cannot justify the reason for the omission of the role of land cover in the hydrological changes of this area on the basis of any published work, nor justify the validity of this omission on the modelling results which are largely inconsistent in this respect.

- Following Bieger et al (2017), the authors can claim that SWAT is "one of the most widely used hydrologic models in the world" but cannot claim that it is a physically-based model.

The model parameters were optimized to obtain the best simulation of discharges, but not the various 'hydrological indicators' extensively exposed in the manuscript. Therefore, due to the equifinality problem (Beven, 2006; Kirchner, 2006), various sets of model parameter may give discharge efficiencies very similar to the best one, but may give quite divergent values of these 'hydrological indicators'.

Consequently, model-simulated 'hydrological indicators' face to two severe uncertainties: the role of land cover changes and the issue of model equifinality. It is not possible to determine whether the differences between modelled and observed trends of these indicators are due to the role of land cover change or modelling equifinality effects. Therefore, the statement in line 76 of the Supplement regarding trend analyses is not acceptable: "Moreover, the fact that streamflow was first calibrated ensures overall the validity of the analysis".

- Throughout the manuscript, the trends are shown as "Sen slopes", but the units are not always shown, especially for the time variable, so the value of the rate of change is not clear if it is per day, month or year.

Removing the stronger influence of river segments with higher streamflow in the analysis of temporal trends does not seem to me a sound option when the main objective of the study concerns water resources. On the other hand, the comparison between slopes and mean flows (fractions of runoff gained or loss at annual intervals) are convenient to evaluate their importance in terms of water resources.

- The hydrographs shown in figures 2 and S3 to S8 are very difficult to understand because the diverse plots cover each other. A logarithmic scale of the discharge axis might help to better visualize the plots.

- The units for the variables x and y in the equation enclosed in Figure 3 are not stated. This graph seems to mix observed and simulated results, which should be explained.

- The analysis of trends in model residuals in section 1 of the Supplement does not justify the validity of excluding land-cover changes in the study, but demonstrates the difficulty of the modelling exercise to provide reliable estimates of the trends.

- The units of the slopes shown in table S1 are not shown. Both this table and figures S1 and S2 demonstrate the very inconsistent results of the model exercise with respect to these trends. These results must be further discussed and the sentence "we can reasonably assume that land use changes in our study are not a main driver influencing streamflow" should be deleted.

- The Y-axes in figures S9 and S10 are not appropriate because variables of diverse ranks are shown together. The ratio of slope to mean value (%) could be better as Y-axe units, as this could increase the visibility of low values and allow direct comparison between gauging stations because the axes could be equal.