Reviewer Comments 1 (RC1)

"The abstract section could be concise".
Answer: We modified abstract. We agree with this comment and corrected the abstract accordingly (see new text).

[General comment], RC1 - "One of the main reasons for the discrepancies between monitoring data/existing literatures and model simulations might be the weakness of SWAT model to capture extreme flows or water yields."
Answer: The authors would like to thank the comments from Reviewer #1 and welcome them. We agree with this general comment. Other detailed responses are described below, as follows.

Original text, Line 55, RC1 – "Colombia (2015, 2014, 2010)"
Answer: Corrected in the updated version of the manuscript. Thank you.

Original text, "Lines 58 to 66, RC1 - "Hoekstra et al., 2011" is over cited. Could be rephrased."
Answer: This entire paragraph was rephrased. We cited Hoekstra et al. (2001) less times.

Line 141, RC1 –"... run from 2009 to 2014".
Answer: Thank you. The new statement is: "The Water Producer/PCJ Project was developed from 2009 to 2014 in the Cantareira System region (Guimarães, 2013), using local actions adopting the concept of Payment for Ecosystem Services-Water [Pagiola et al, 2013; quoted] ".

Line 153, RC1 - "three data collection platforms" their geographic locations could be indicated on the study area map."
Answer: The three DCPs are shown in the study area (in Table 1, Table 4 and Figure 8, in the new version of the manuscript).

Line 156, RC1: "the type of secondary data could be clearly indicated."
Answer: We appreciate this comment. The explanation to be updated in the new version of the paper appears as follows (because of the extension of these new statements, we suggest including them in a Supplementary Material section, according to the HESS Editor´s final decision): "To reduce uncertainty about hydrological scaling effects of EbA through LULC scenarios from 2011 to 2014, we also collected supplementary, secondary data using three strategies. First, we scheduled field observations and interviews with local landowners and farmers who explained their past, present and future (planned) best management practices related to Payment for Ecosystem Services-Water, derived from EbA initiatives from the *PCJ-Produtor de Agua* Project of the Cantareira System´s headwaters [Pagiola et al, 2013, Brazil's Experience with Payments for Environmental Services. Payments for Environmental Services (PES) learning paper; no. 2013-1. World Bank, Washington, DC, World Bank. https://openknowledge.worldbank.org/handle/10986/17854 License: CC BY 3.0 IGO]. This secondary information helped to link LULC derived from EbA/PES-Water with some parameters of selected hydrologic response units (i.e. SWAT-HRUs). These field observations on the local knowledge brought a better understanding about physically-based parameters calibrated regionally, but with unsatisfatory coefficients in some

catchments, i.e. Posses Catchment (13-km^2 drainage area). Second, we also collected secondary information about the stakeholders´ opinions concerning the 23 scenarios we developed in this paper from the multi-agent, multi-level governance of the *PCJ-Produtor de Agua* Project (municipality, state and national). Due to the states' border between Minas Gerais (MG) and São Paulo (SP), which have different reference standards, the various stakeholders´ opinions strongly influenced PES-Water/EbA practices across the transboundary (inter-state) nature of most Cantareira System's catchments. Thus, we undertook extra field visits to evaluate sites with the greatest uncertainty in modelling EbA and LULC scenarios to receive new flow gauging stations. These stations were selected together with representative decision-makers from the states and municipalities that are part of the sub-basins studied (Extrema-MG, Joanópolis-SP, Piracaia-SP and Nazaré Paulista-SP), states (IGAM-MG, SMA-SP and DAEE-SP), federal agencies (ANA-The Brazilian Water Agency, CPRM- Brazilian Geologic Survey, and the National Center for Monitoring & Alerts of Disasters, CEMADEN-MCTIC) and non-government organizations (WWF-Brazil, TNC-Brazil and local initiatives) (see Taffarello et al (2016-b), http://dx.doi.org/10.4236/jep.2016.712152). Third, the aforementioned strategies helped identify, select and prioritize qualitative and quantitative variables to reduce the uncertainties in the generation of pollutant loads under LULC, as proposed by other authors (see e.g. Zaffani et al, 2015; doi:10.4172/2161-0398.1000173, quoted in the references). These secondary data revealed the most viable conditions for nested catchment experiments to monitor and test hypotheses through a scenario-intercomparison modelling of upstream areas of the Jaguari-Jacareí, Cachoeira and Atibainha reservoirs, and are updated regularly by official agencies with open access repositories of hydrological databases, such as ANA (http://hydroweb.ana.gov.br) and CEMADEN (http://www.cemaden.gov.br/pluviometros-automatico/)".

Lines 252-255, RC1: "Besides adopting from the existing literatures, implementing sensitivity analysis could be recommended in order to select model parameters."
Answer: We appreciate this comment. The explanation to be updated in the new version of the paper appears as follows (because of adding these new statements, we suggest including them in a Supplementary Material section, according to the HESS Editor´s final decision): "Modelling parameters for water yield calibration were selected not only by consulting the SWAT literature [i.e. Arnold et al, 2012; Bressiani et al, 2015; Fukunaga et al, 2015; Gassman et al, 2007; see more explanations for other review comments below], but also by performing supervised analysis and comparing parameters from recent literature [i.e. Francesconi, W., R. Srinivasan, E. Pérez-Miñana, S.P. Willcock, M. Quintero. 2016]. "Using the Soil and Water Assessment Tool (SWAT) to model ecosystem services: A systematic review", *Journal of Hydrology* 535 (2016) 625–636. DOI:10.1016/j.jhydrol.2016.01.034, and Monteiro, J. A. F., Kamali, B., Srinivasan, R., Abbaspour, K., and Gücker, B. (2016) "Modelling the effect of riparian vegetation restoration on sediment transport in a human-impacted Brazilian catchment", Ecohydrol., doi: 10.1002/eco.1726, now quoted] and even from consulting the USP open access repository [see studies by Rodrigues, 2014, www.teses.usp.br/teses/disponiveis/18/18138/tde-18122014-094354/pt-br.php; Bressiani, 2016, www.teses.usp.br/teses/disponiveis/18/18138/tde-04042017-155701/pt-br.php, and Mohor, 2016, www.teses.usp.br/teses/disponiveis/18/18138/tde-23032017-102949/pt-br.php]. First, in spite of a much larger list of suggested parameters for modelling goals proposed by Bressiani (2016; quoted), our regional sensitivity analysis followed the recommendations of the theory and practice of mapping ecosystem services

using Tier 1 and Tier 2 models [see Mendoza et al, 2012, Ch. 3, in Kareiva et al(eds), 2012; ISBN 978-0-19-958899-2] constrained by the short time series monitored for all sites, with inequal quantitative assessment, seasonality and scale effects. Second, based on the studies carried out by Rodrigues et al [2014, doi:10.1002/2013WR014274, 2015, doi: 10.1002/2014WR016691], Bressiani et al [2015, doi: 10.3965/j.ijabe.20150803.1765] and Mohor & Mendiondo [2017, doi: 10.1016/j.ecolecon.2017.04.014], we selected 18 SWAT parameters and their initial range of combinations, as follows: Available water capacity, Moist bulk density, Saturated hydraulic conductivity, Baseflow alpha factor, GW-Revap Coefficient, Groundwater delay time, Deep aquifer percolation fraction, Threshold depth of water in the shallow aquifer for "revap" or percolation to the deep aquifer to occur, Soil evaporation compensation factor, Plant uptake compensation factor, Manning´s roughness for the main channel, Effective hydraulic conductivity in the main channel, Maximum canopy storage, Manning´s for overland flow, Average slope steepness, Initial SCS CN (for antecedent moisture condition 2), and Surface runoff lag coefficient. Thirdly, in-situ field validation tests were developed through experimental campaigns to test the limits of variation of streamflow and water quality (see explantations below).

Lines 276 to 279, RC1 - "It is known that SWAT model is not for extreme flows and hence water quality parameters."
Answer: We agree with this comment. For EbA scenario purposes, we planned to set up field investigations and SWAT calibrations [see Figure 5, HESSD paper] using the extreme conditions of 2013–14 drought through quali-quantitative freshwater monitoring at the headwaters of the Cantareira System, quoted in this paper [see i.e. Tafarello et al, 2016; doi: 10.1080/02508060.2016.1188352]. This evidence outlined water quality results from 17 catchments, showing regional behaviour for water quality loads in drainage areas (ranging 0.66–925 km^2) for future modelling parameterization through SWAT for EbA scenarios purposes. We experimentally sampled water quality parameters of pH, water temperature, electrical conductivity, turbidity, biological oxygen demand (BOD), chemical oxygen demand (COD), total solids (TS), NO3, NO 2, PO 4, thermotolerant coliforms and Escherichia coli, in several catchments, varying the drainage area, the land use and land cover, which helped us to address the uncertainty and complexity of factors affecting SWAT parameter selection. Moreover, a summary of these results are detailed in Table 4 (this HESSD paper).

Line 299 (RC1) could be moved to line 298.
Answer: It was corrected in the updated version of the manuscript. Thank you.

Lines 310 (RC1) could be moved to line 309.
Answer: It was corrected in the updated version of the manuscript. Thank you.

Lines 322 (RC1) could be moved to line 321.
Answer: It was corrected in the updated version of the manuscript.

Lines 455 to 456 (RC1) should be written with appropriate multiplication sign.
Answer: Rewritten as: "... The 52% decrease of water yield between S1 (1990) and S2 (2010) scenarios, as (14.9 -31.3)/31.3×100) might be related to a marginal increase in the Eucalyptus cover..."

Line 514, RC1: "It would be useful to relate spatially the sub-basins in which the differences in land-use/land-cover are the greatest and the water yield, nitrate, total phosphorus and sediments yield differences are evident. For instance providing maps which indicate temporal changes in LULC and corresponding changes in water quality parameters considered."

Answer: We appreciate this comment. The explanation to be updated in the new version of the paper appears as follows (because of adding these new statements, we suggest including them in a Supplementary Material section, according to the HESS Editor´s final decision): "Due to the significant variabilities among selected basins where in-situ monitoring was developed for EbA scenario purposes, and because we have not performed field validation in all distributed HRU (hydrologic response units), we decided not to show the regional results through maps. Whatever interpolation technique used, it will not be able to catch the inherent ground-context heteregeneity, and physically-based characteristics of high-variability functionality of these subtropical catchments. Instead, we performed an initial analysis of clustering similar responses from catchments with the most plausible explanations as follows. On the one hand, evidence of SWAT modelled scenarios showed two groups of river basins under EbA scenarios, with distinct land use change of native forest fractions (NF%). Our results show Group 1, with 11 of the studied basins, with native forest recovery using EbA (S2+EbA), as well as an intermediate land use fraction as follows: NF%(S2)< NF%(S2+EbA)<NF%(S1). In turn, Group 2 of 9 river basins showed a progressive growing fraction of native forests across scenarios, with best EbA land use impacts, as follows: NF%(S1)< NF%(S2)<NF%(S2+EbA). The basins from Group 1 are mainly located close to both urban settlements and Eucaliptus plantations in Northwestern headwaters, where conservation projects have minimum adherence and no significant effect on LULC and, therefore, on SWAT outputs (see Figure 3). Moreover, the catchments from Group 1 are mainly located in Eastern and Southeastern areas (Figure 3), where there are more EbA projects of *PCJ-Produtor de Agua*. On the other hand, the greatest impacts in water yield are inversely correlated with land-uses and water pollutant quality, but with high non-linear relationships and without explicit regional factors (see Figure 11). For an integrated assessment of hydro-services, it is worth noting that phosphorus, nitrate and sediment yields have spatio-temporal changes of load production across scenarios S1, S2 and S2+EbA, which would be better understood in selected catchments, namely Alto Jaguari and Domithildes.

Line 533, RC1: "Reason for selecting the two sub-basins among the 20 sub-catchments?"
Answer: We appreciate this comment. The explanation to be updated in the new version of the paper appears as follows (because of adding these new statements, we suggest including them in a Supplementary Material section, according to the HESS Editor´s final decision): "These two catchments were selected regarding the different groups identified in this study, contrasting the outputs from 20 sites: Upper Jaguari is selected from Group 1 and Domithildes is selected from Group 2 [see comment 12]. Moreover, we studied the following variables in the two selected catchments. First, we analysed the fraction of water yield affected by the grey water footprint for nitrate (ca. 0.08 to 3.9 mg/L), total phosphorous (from 0.02 to 1.2 mg/L) and sediments (approx. 0.03 to 250 mg/L). These concentrations represented dilution demands between 0.1 % to close to 1000 % of simulated water yield for a wide range, in between 10 to 500 km^2 [see Figure 12, this HESSD paper]. Second, these demands depended on: the native forest cover [i.e. in Figure 9, with S1 for year 1990, S2 for year 2010 and S2+EbA for year 2035], the flow duration curves under three LULC scenarios at 20 headwaters [Fig. 10], and the scaling effects of

EbA actions on drainage areas [ranging from the small Domithildes catchment of 9.9 km^2 to the medium-sized Alto Jaguari catchment of 302 km^2]. These factors clearly affected (a) the fraction of water yield affected by the GWF-NO$_3$, GWF-TP and GWF-Sed, and (b) the reference flows in duration curves, both in streamflow and in pollutant loads, especially for low-flows (higher duration probabilities [see Fig. 13 and 14]. Moreover, the annual regime of water yield of these two selected catchments revealed local constraints in the size of catchments ranging from 10 to 300 km^2. Thus, we pointed out the limits for SWAT modeling when using the EbA assessment and PES-Water projects, by using the grey water footprint, ranging from GWF-NO3 below 0.2 m^3/s to GWF-TP up to 20 m^3/s. These results converged with the general discussion with blue and green water accounting shown in the studies carried out by Rodrigues et al [2014; A blue/green water-based accounting framework for assessment of water security, Water Resour. Res., 50, 7187–7205, doi:10.1002/2013WR014274], now quoted in the references of this manuscript.

Line 535, RC1: "Any statistical relationship between the changes in LULC classes and grey water footprints. For instance multivariate statistical analysis."

Answer: For instance, the evidence we modelled using SWAT concerning GWF and LULC was presented in lines 514 to 534 (first version of the manuscript). These results refer to regionally average values (20 catchments), the same test period (8-yr time series tested) and with the fixed time-step modelled (SWAT monthly-basis). On the one hand, native forest land use fractions (NF%) have ranges of 41±14, 39±15% and 44±16 %, and were related to GWF-NO3 of 0.68±0.6, 0.28±0.1, and 0.44±0.1, for S1(1990), S2(2010) and S2+EbA(2035) scenarios, respectively. On the other hand, medium-sized vegetation land use fraction (native, eucaliptus and orchard) ranged between 46%, 53% and 62% for the same scenarios, respectively, not showing a trend. For GWF-TP and GWF-Sed, the values differ in absolute terms and the averaged ratios of GWF/Water Yield also changed. In spite of the high variability of responses, and small period of testing, we recommend future field campaigns and further multivariate statistical analysis, but they are out of the scope of the present manuscript.

Lines 544 to 555, RC1: "As one-third of the SWAT simulation are low-flow or drought years. It is known that SWAT model is weak in capturing extreme flows. One of the reasons for the discrepancy between monitoring data and model simulation might not the weakness of the SWAT model to represent low-flows?"

Answer: We agree with these comments. On the one hand, recent papers addressing a review of SWAT applications in Brazil outlined the challenges and prospects to reduce the discrepancies between monitoring data and existing (regional) literature and model simulations [i.e. Bressiani et al, 2015; doi: 10.3965/j.ijabe.20150803.1765], quoted in the references. This general review is useful to address model discrepancies in a multilevel approach: quantitative water yield, water quality loads and rainfall-streamflow behaviours at a range of scales during the same period of monitoring and the inherent streamflow variability at these subtropical catchments. Due to this, our strategy selected sites through a nested catchment experiment to study these discrepancies according to the natural hydrological cycle, when possible. On the other hand, we addressed these discrepancies by quantitative calibration with a consecutive freshwater quality calibration. Our evidence showed [see i.e. Fig. 5] that at some drainage areas, between 12 km^2 to 508 km^2, the SWAT model might underestimate observed streamflows. In three out of four campaigns, the results of both flow rates and nitrate loading (NO3) were very

close to the values simulated by SWAT. Only the campaign conducted in May, 2014 demonstrated a significant difference between field validation with SWAT modelling, which may have occurred because of the SWAT limitation in updating loads (water quality parameters) with more prolonged dry periods, as discussed in the papers by Taffarello et al [2016-a; doi: 10.1080/02508060.2016.1188352] and Mohor & Mendiondo (2017; doi: 10.1016/j.ecolecon.2017.04.014, quoted].

Table 1, RC1: "It might be better to replace sub-basin coordinates with key modelling results and/or field observations."
Answer: In the new, updated manuscript, we included new columns, pointing out modelling results and field observations.

Table 2, RC1: "Possible reason for model underperformance for some sub-basins?"
Answer: As mentioned in the paper, both the Posses catchment and Cachoeira catchment have been constrained by limitations in SWAT modeling set-ups because of: anthropic and ilegal domestic water withdrawals across riversides and margins, with small dams affecting the streamflow regime and in some cases, Eucaliptus sp planted close to river channel during low-flows. Taffarello [2016, quoted] showed this in the open-access repository pictures, which described anthropic impacts on water yield and water withdrawal [see www.teses.usp.br/teses/disponiveis/18/18138/tde-05042017-091421/pt-br.php]. These human-made impacts strongly affected the SWAT underperformance in calibration and validation steps, not only on NASH, NASH-log but also on the PBIAS, especially after long periods of droughts or rainfall anomalies [see Figure 5, this HESSD paper]. Because these human-made interferences come from real situations at the catchments studied, without special SWAT parameterisation and scaling from HRU to the whole catchments, we decided not to reduce both complexity and heteregeneity through a complete, exhaustive sensitivity analysis of SWAT parameters. Instead, we recommend further studies along these lines if new and more field evidence in other catchments is made available.

Table 3, RC1: "The selected SWAT parameters are not exhaustive unless sensitivity analysis is conducted."
Answer: The main objective of this paper submitted to HESS is not to address sensitivity analysis among SWAT parameters. Instead, we aimed to perform hypothesis tests of scenario intercomparisons, including EbA policies and PES-Water actions, using SWAT pre-calibrated parameters, linked with previous field evidence collected during sampling periods and previous modelling experiences in these basins [i.e. Rodrigues et al, 2014, doi: 10.1002/2013WR014274; Rodrigues et al, 2015; Bressiani et al, 2015, DOI: 10.1002/2014WR016691; Taffarello et al, 2016-a, DOI:10.1080/02508060.2016.1188352; Mohor & Mendiondo, 2017, DOI: 10.1016/j.ecolecon.2017.04.014]. It is worth noting that this paper submitted to HESS is one of the first Brazilian contributions of coupling EbA directives into hydrological modelling using nested catchment experiments in the Brazilian Atlantic Forest [see Taffarello et al, 2016-b, DOI: 10.1016/j.cliser.2017.10.005], promoting other research groups which might develop furhter modelling hypotheses. Regarding the sensitivity analysis, we proceeded in the calibration process, although it was not exhaustive. On the other hand, and given that SWAT has a very large number of parameters and our

experiment involved nested catchments, rather than a single experimental basin, testing all parameters in our study case with EbA would be rather laborious. As mentioned earlier, we consulted previous applications of SWAT in the literature, preferably those in Brazilian basins, to find the most indicated parameters to work on. Based on Fukunaga et al. [2015, DOI: 10.1016/j.catena.2014.10.032], Gassman et al. [2007, DOI: 10.13031/2013.23637], Arnold et al. [2012, DOI: 10.13031/2013.42256] and a good review by Bressiani et al. [2015, quoted], we first selected 18 SWAT parameters with their initial ranges by Rodrigues et al [2014, 2015 quoted]. Then, we made analyses of these 18 parameters in our sub-basins. After analyzing these results, we chose to re-calibrate parameters in some basins. Thus, SWAT-CUP was performed in our tests, with each cycle consisting of 300 runs. In each cycle, we reached new limits for each parameter or stopped tuning a parameter. The number of cycles varied among the sub-basins, from one to five cycles. From all the 20 nested catchments studied, and using the initial 18 SWAT parameters, some sites completed the calibration with 7 calibrated parameters, while others had a total of 17 - out of those initial 18 parameters. From upstream to downstream, after the automatic step, a manual calibration refinement also took place. One example of the range of the final values is shown below in Table A.1 (new).

Table A.1: Range of coefficients adopted for calibration in SWAT-CUP and final values found after manual stage calibration.

Parameter*	Initial (mín	Median[1] mín	Chosen mín	Chosen máx	Median[1] máx	Initial máx
a__CANMX.hru	0	0	0	100	60	100
a__Ch_N2.rte	-0.0005	0	0	0.28	0.3	0.3
a__CN2.mgt	-15	-12	-8.67	10.31	10	15
a__GW_DELAY.gw	-15	-3	-4.161	42.69	30	50
a__GWQMN.gw	-550	-300	-415.02	360.00	350	450
r__OV_N.hru	-0.5	dismissed	-		dismissed	1
r__SHALLST.gw	-0.5	-0.3	-0.08	0.39	0.4	0.6
r__SOL_AWC().sol	-0.5	-0.25	-0.42	0.29	0.33	0.5
r__SOL_BD(1).sol	-0.2	-0.15	-0.19	0.18	0.2	0.4
r__SOL_K().sol	-0.4	-0.27	-0.32	0.35	0.37	0.5
v__Alpha_BF.gw	0.01	0.02	0.001	0.049	0.05	0.1
v__Ch_K2.rte	0	0	0	36.74	30	130
v__EPCO.hru	0.4	0.4	0.85	-[2]	1	1
v__ESCO.hru	0.4	0.7	0.69	0.95	0.95	0.95
v__GW_REVAP.gw	0.02	0.02	0.01	0.18	0.2	0.2
v__RCHRG_DP.gw	0.01	0.01	0.05	0.68	0.5	1
v__REVAPMN.gw	0	500	539.28	959.28	1000	1000
v__SURLAG.hru	0.01	1.5	0.97	5.53	4	5
IPET			(0) Priestley-Taylor			

Legends: "1": "median" of the limits adopted in following runs in SWAT-CUP. Manual calibration could overcome these limits; "2": only one sub-basin had EPCO modified. * a_ stands for "added" value, i.e. the final value in each feature (e.g. each HRU) is the original value plus the calibrated coefficient; r_ stands for ratio, i.e. the final value in each

feature is the original value times 1+ the calibrated coefficient; v_ stands for value, i.e. the final value of the feature is the calibrated coefficient.

References cited:

Arnold, J. G.; Moriasi, D. N.; Gassman, P. W.; Abbaspour, K. C.; White, M. J.; Srinivasan, R. et al. (2012): SWAT. Model Use, Calibration, and Validation. Em: *Transactions of the ASABE* 55 (4), pág. 1491–1508. DOI: 10.13031/2013.42256.

Bressiani D A, Gassman P W, Fernandes J G, Garbossa L H P, Srinivasan R, Bonumá N B, et al. (2015): Review of Soil and Water Assessment Tool (SWAT) applications in Brazil: Challenges and prospects. Em: *Int J Agric & Biol Eng,* 8(3), pág. 9–35. DOI: 10.3965/j.ijabe.20150803.1765.

Fukunaga, Danilo Costa; Cecílio, Roberto Avelino; Zanetti, Sidney Sára; Oliveira, Laís Thomazini; Caiado, Marco Aurélio Costa (2015): Application of the SWAT hydrologic model to a tropical watershed at Brazil. Em: *CATENA* 125, pág. 206–213. DOI: 10.1016/j.catena.2014.10.032.

Gassman, P. W.; Reyes, M. R.; Green, C. H.; Arnold, J. G. (2007): The Soil and Water Assessment Tool. Historical Development, Applications, and Future Research Directions. Em: *Transactions of the ASABE* 50 (4), pág. 1211–1250. DOI: 10.13031/2013.23637.

Rodrigues, D.B.B.; Gupta, H.V.; Mendiondo, E. M. (2015): Assessing uncertainties in surface water security: An empirical multimodel approach. *Water Resources Research*, 51 (11): 9013–9028. DOI: 10.1002/2014WR016691.

Rodrigues, D.B.B.; Gupta, H.V.; Mendiondo, E. M. (2014): A blue/green water-based accounting framework for assessment of water security. *Water Resources Research*, 50 (9): 7187–7205. DOI: 10.1002/2013WR014274.

Table 5, RC1: "I would like to see additional column indicating the Hydrologic Services Index. The symbol used for the sub-basins 10, 15, 17 and 19 is not defined."
Answer: We answered this comment, including this HSI value as a new column. The symbol used for sub-basins 10, 15, 17 and 19 was a typing error. We appreciate your comment. Thank you.

Figure 2, RC1: "Sensitivity analysis is missing after SWAT-CUP".
Answer: In this manuscript, as mentioned, we followed a step-by-step, but not exhaustive, calibration procedure using collection and assessment of data, understanding the watersheds, identifying and selecting sites and periods to calibrate and validate, defining calibration methods, objective functions and evaluation metrics, main water balance components, with volumes and process representations, defining parameters and ranges of variability, sensitivity analysis, calibration, validation, cross validation and uncertainty analysis (Bressiani, 2016, quoted; Mohor, 2016, quoted). As previously mentioned, we also consulted former SWAT modelling strategies used in these basins,

available in the open repository by Mohor [2016; www.teses.usp.br/teses/disponiveis/18/18138/tde-23032017-102949/pt-br.php], Bressiani [2016; www.teses.usp.br/teses/disponiveis/18/18138/tde-04042017-155701/pt-br.php] and Rodrigues [2014; http://www.teses.usp.br/teses/disponiveis/18/18138/tde-18122014-094354/pt-br.php]. In our paper, we addressed a calibration stage of SWAT-CUP (Calibration and Uncertainty Programs) software and SUFI-2 (Sequential Uncertainty Fitting) method. SUFI-2 is based on Latin Hypercube sampling [Abbaspour et al, 2015; quoted in the references). After this automatic stage, a finer adjustment using manual calibration was made, following the recommendations of Mohor (2016) and Mohor & Mendiondo (2017; DOI: 10.1016/j.ecolecon.2017.04.014), quoted in the references. For more in-depth sensitivity analysis of SWAT parameters, we recommend Bressiani (2016) who proposed not only a new systematic procedure for calibrating the SWAT model in complex basins, but also a search for a better SWAT performance and reduced optimization time, using different calibration methods on different watershed locations. Moreover, Rodrigues [2014, Table 2.3, page 56] adjusted some parameters for nested catchments in the Cantareira System (CN2, Canmx, OV_N, SOL_K, SOL_AWC), according to land use classes.

Figure 4, RC1: "Why the upper and lower bound of coef. of PBIAS is only ± 0.15, though the model performance for some sub-basins are more than ± 0.15."
Answer: We appreciate your comment, thank you. We corrected Figure 04.

Figure 6, RC1: "How representative is the sampling of only 8 months for turbidity?"
Answer: During the 2013/2014 field campaigns across all the nested catchments presented here, the turbidity ranged between extremes of 1 and 300 NTU, with median values close to 11 NTU. These high variability captured ranges of in-situ monitored instantanous mean cross-section velocities below 1 m/s and specific streamflows ca. 0.001 to 0.025 m^3/s/km^2. These values captured approximate flow discharges in the range of 5% and 96% of probability of regional flow duration curves, and also affected the variability of the turbidity of water quality. Moreover, these ranges were observed during the 2013/2014 anomalous rainy season, alternating heavy rains and dry periods, in both reference catchments with EbA initiatives and impacted catchments with land-use changes. Due to this, we understand, in spite of having sampled only 8 months of monitoring, observed turbidity is not biased and could represent the conditions for using EbA hypothesis for the scenarios we tested. More details of experimental sampling and observational schemes are explained in Taffarello et al [2016-a; quoted].

Figure 12-a, RC1: "Legend for y-axis has typo error."
Answer We appreciate your comment. It was corrected in the updated version Fig. 12-a.

Figures 13, 14 and 17, RC1: "The legends and axis values are not readable."
Answer: The legends of Figures 13, 14 and 17 were increased in size and readability is now much better. We appreciate your feedback concerning this correction.