Review ESSD-2017-118

Somewhat surprised to see this paper submitted to ESSD. But a good direction for the journal and compliments to the authors for attempting to make their data publicly accessible. I share the authors view that they address a critical region of our planet.

Fundamentally I regard the data gathering, species identification and data analysis as sound. As this review will show I have only a few comments and suggestions on those aspects. I believe however that the data and the paper need a stronger and clearer presentation to appeal to a wide audience. On those presentation aspects I have made many comments and suggestions, enough perhaps to require a substantial revision. If the editors agree I will suggest a major revision that should lead to eventual publication.

The Pangaea link works very well. The authors basically repeat the tables from the manuscript as data files in the Pangaea archive. Because the Pangaea landing page correctly identifies the four data tables (tables 1, 3, 5 and 6) as tab-delimited files, the names of the files in the dataset folder should carry this designation. E.g. Arunkumar-etal_2015-T3.tsv rather than Arunkumar-etal_2015-T3.tab. R, Excel and other spreadsheets can easily ingest a .tsv file but those software packages will not recognise a .tab file. Once I renamed them I had no problem to open all the .tsv files.

Lines 13, 14. This sentence at the start of the abstract, about studying the freshwater fish from 2010 to 2013 can give a wrong impression. Some readers will assume that you studied these rivers in repetitive years, e.g. in 2010, again in 2011, out to 2013. In fact this study reports the outcome of a collection and identification process that covered 31 separate sampling sites that required 4 years (2010 to 2013) to complete. Nowhere in the documentation do we read about any repeat sampling. You should make very clear that you conducted one comprehensive sampling and assessment of each site in a process that required 4 years, 2010 to 2013.

Lines 16, 17. These lines, about 64 species, some many orders, families, etc. repeat information from line 14. Remove the sentence in line 14? We do not need to see this information twice, so close together.

Line 18. Plural of ‘genus’ should appear as ‘genera’ (as in line 17 above)?

Lines 23 to 26. The collection and identification of some fish species considered endangered or critically endangered probably represents an import contribution of this study. These fragmented and confused statements do not provide an adequate summary. The manuscript that follows probably needs an explicit section on endangered species and the significance of finding them in the Western Ghats. This abstract implies a “concise discussion” but that discussion never appears in the subsequent manuscript?

The introductory paragraphs look good.

Line 45, 46. The “Satpura” hypothesis? Specific to Indian or South Asian ecosystems or something more general. How and why did that hypothesis stimulate discussion on, for example, “endemism”. Does this discussion have relevance to this paper?

Lines 48 to 62, This paragraph about new species discoveries has relevance to the larger point about the Western Ghats as a location and source of unique biodiversity? We need to know the reason for this list of species names.

Line 65, 66. This combination of various types of nets has some minimum capture size? Larger organisms could avoid the nets? Please give users a sense of the size range captured by your sampling. What happens to any invertebrates? Discarded, recorded, ignored, or not captured?

Line 67. Five specimens from each species. This means you did some species identification in the field? Did you have a statistical basis for this sampling strategy? Or a valid logistical constraint? By focusing on repeatable numbers of species present, you have minimised the presence-absence question? The issue of absence of expected species does not arise in this manuscript, despite earlier mention of disturbance and invasives? Perhaps this discussion belongs in a subsequent research paper but those future researchers will need to know how to understand your data. Absence means zero specimens of a given species collected or fewer than 5 specimens collected?

Line 81, water samples collected “post-monsoon”. We should have actual dates for all collection episodes, fish and water? Perhaps these exist in your database but they do not appear in any of the data shared here.

Lines 98, 99. Confusion here for the reader about primary sources of uncertainty. One source, identified here, involves conversion of information from “elementary” or “original” data sheets? Which feeds which? Which represents inputs to the database? Errors frequent or rare in these data translation processes? We need more clarification here. Later, in the summary, you should list for users all the known sources of uncertainty.

Line 104. Present collection sites based on prior literature reports? Shouldn’t we have this information, along with discussion of impact on reliability and repeatability, earlier, in the methods section? If true, this opens the possibility of comparison of abundance, species presence-absence, etc., with earlier collections? We need to know if, where, and to what degree the present collections enable these historical comparisons. We do not need the comparisons themselves - those perhaps belong in a separate research paper - but we do need to know if the current collections enable such comparisons with prior collections. If not, why not?

Lines 104 to 112. This description of geomorphology and biodiversity of the Western Ghats belongs in the introduction?

Line 114. We need a much better presentation of the sampling sites. Figure 1 shows only 16 out of 31 sites. If the authors want to show locations on a map, we need a better, more complete map. In Table 1 we get very useful information of sites by river system with elevation and lat lon for each site. But in Table 5 and 6 we lose that information. Those tables should still make clear the association of sites with rivers; you do not want us as users making mistakes in our assignment of site to river system. Table 1 does not seem to sort the site by elevation; somewhere we need that information. In the .tsv data file for table 1 I could sort by river system and then elevation (or forest type); better the authors should do this for all readers?

Line 119. Discussion of diversity, abundance and distribution by site, but not by river system, elevation, drainage area, proximity to human influence, presence or absence of hydroelectric dams, etc. Later in this results section the authors mention elevation, water temperature, water quality, lakes, etc. But here we don’t get any sense of those factors for all sites or for each river system; we only see site-by-site lists.

Line 128. Species similarity was very ‘low’ rather than very “less”? The authors only hint at all the factors that might impact similarity, e.g. as plotted in Figure 6. Again we would need to see those sites according to their river systems rather than independently? Or, do the authors imply that site-to-site differences exceed river-to-river differences? We do not get a clear treatment of river vs site data and differences that would allow us to assess the validity of such similarities or differences. A clearer presentation of sites by river system, and of summary statistics of river vs river, or of the absence of significant differences, would help!

Line 138. IS: 10500 Permissible limits. This represents an India-wide water quality standard? We need a reference to it?

Lines 137 to 154. This water quality discussion occurs almost entirely by site, not by river system. Why? Here and in Table 6, I feel surprised to see salinity. Conductivity I expect, and perhaps resistivity, but what do these very low values of ‘salinity’ represent and report? Do we have a definition of the salts involved? Not the typical seawater salts, presumably, so we must have a different freshwater definition of salinity? In other ESSD papers reporting mountain stream water quality, they typically do not report salinity. Because all these values lie below 0.5 ppt (one commonly-accepted definition of freshwater) we should consider them all to have negligible amounts of salts? If the authors do not make explicit use of the salinity data, we do not need them?

Lines 156 to 210. After the water quality paragraph the authors provide four paragraphs on species appearance, diversity, abundance, etc. without clear conclusion and particularly without confirmation of Western Ghats as a biodiversity hotspot or as a ‘refuge’ for endangered species. Either we need less of species lists or we need more synthesis and assessment of the fundamental question: will these data allow and support discussion and conclusion about regional biodiversity and biological refuges or not - even if those discussions occur in other research papers or in other conservation fora. Readers need to get from the authors a sense of confidence on how to use these data! Unfortunately, from this somewhat confused and random discussion of physical and biological influences on habitats and species presence or abundance, we fail to get a clear understanding of the authors’ confidence in their own data. We also learn that the Periyar river flows westward when earlier we read that this study focused on eastward-flowing rivers and about the existence of Periyar Lake. Certainly the map in Figure 1 and none of the text so far gave us any hints of lakes.

Line 188: “moolavaigae”? Presumably this refers to a high-elevation location of Moolavaigae?

Line 212, 213, Summary. I do not understand the point of the first sentence about morphological variation related to micro- or macro-habitat?

Lines 220 to 222. Again by individual sites, with no reference to river systems.

Line 223: The present study failed to convince this reader that altitude had any consistent effect. Probably more related to a weakness in the presentation rather than a weakness in the data, but in either case the paper has not shown us convincing data relating biodiversity to elevation.

Lines 224 to 232: Perhaps valid statements toward a positive “ecological spirit”, but in fact from these data the authors have not guided us to conclusions about “sharp decline” (line 227) or about social pressures.

A good conclusion should briefly summarise the data, explicitly caution users about uncertainties and limitations of the data, and then outline both the present impacts (for conservation management) and need or intention for future monitoring or data gathering.

Table 1. ‘Forest Type’: these terms come from FAO or GBIF definitions or from an India Forest Classification scheme? Readers need to know how to relate this terminology to other data from other regions. ‘Stream Order’: hydrologists will understand this general mechanism to indicate stream branching but the authors should specify whether these represent standard Strahler stream order numbers or some other India-specific index of stream branching? ‘Area’ presumably represent catchment area above (upstream) of the sampling location. The authors should inform readers how they calculated this or from what source they extracted this information. Likewise for ‘Volume’, this presumably represents annual mean volume measured at some exit or drainage point of each stream and river? Again, readers need to know where this information comes from. In this table, ‘Mean Velocity’ represents a code referenced in a footnote, and not an absolute value? The authors could reduce confusion by using the actual codes ‘slow’, ‘moderate’, ‘very fast’, etc?

Table 2. Comprehensive species list but the IUCN codes in column 4 do NOT match the descriptions in the footnote. For example, the footnote does not define LC as it appears in the Table while ‘LRnt’ from the footnote never appears in the Table.

Table 3. We need these data explicitly organised by river system and perhaps sorted in order of elevation within each river system. With effort a user can establish these distinctions and filter by elevation by using the .tsv file but the authors need these improvements in Table 3 in order to support their discussion about, for example, biodiversity and altitude?

Table 5. Species by generic number code vs. site by similar generic number code. Inclusion of actual species names (as genus.sp) and clearer organisation by river system would greatly increase the utility and information content of this Table!

Table 6. Again, organise this by river system and sort by elevation?

Figure 1. The authors should cite their source for the base map? Not particularly useful as presented because it includes only about half of all sampling sites, gives no highlight of selected river systems, shows no lakes or dams, etc. Presumably a DEM exists for this region but perhaps not at the resolution needed? Many conservation organisations have better maps? Even on a lower resolution map the authors could label their sampling sites while also emphasising the biodiversity importance of this region?

Figure 2. Not sure what this figure shows us? Would it offer a basis for comparison to another region in India or to another mountainous biodiverse region elsewhere?

Figures 3, 4 and 5. All these figures need indication of the sampling sites within their respective river system and - if the authors want to focus on elevation - sort within each river by altitude?

Figure 6. We need more information about the numerical basis for similarity-dissimilarity. A user doesn't gain much useful information from this without designation of the rivers? Or perhaps of elevation?

Figure 7. Do these pictures come from this collection or from other sources? I suspect ESSD can not publish them without attribution. Do individual pictures associate with appropriate species in the database?